Adapting extreme value statistics to financial time series: dealing with bias and serial dependence

Laurens de Haan, Cécile Mercadier, Chen Zhou

Adapting extreme value statistics to financial time series: dealing with bias and serial dependence

Laurens DE HAAN[(1)]
ldehaan@ese.eur.nl

[(1)]Dept. of Economics,
Erasmus university,
P.O. Box 1738,
3000 DR Rotterdam,
The Netherlands.

Cécile MERCADIER[(2)]
mercadier@math.univ-lyon1.fr
[(2)]Université de Lyon,
CNRS UMR 5208,
Université Lyon 1,
Institut Camille Jordan,
43 blvd du 11 novembre 1918,
69622 Villeurbanne, France.

Chen ZHOU[(1,3)]
zhou@ese.eur.nl

[(3)]Economic and Research Div.
De Nederlandsche Bank,
P.O. Box 98,
1000AB Amsterdam,
The Netherlands.

Abstract

We handle two major issues in applying extreme value analysis to financial time series, bias and serial dependence, jointly. This is achieved by studying bias correction method when observations exhibit weakly serial dependence, namely the $\beta-$mixing series. For estimating the extreme value index, we propose an asymptotically unbiased estimator and prove its asymptotic normality under the $\beta-$mixing condition. The bias correction procedure and the dependence structure have a joint impact on the asymptotic variance of the estimator. Then, we construct an asymptotically unbiased estimator of high quantiles. We apply the new method to estimate the Value-at-Risk of the daily return on the Dow Jones Industrial Average Index.

Keywords: Hill estimator; bias correction; β-mixing condition; tail empirical quantile function

AMS 2000 Subject Classification: Primary 62G32; secondary 60G70

1 Introduction

In financial risk management, a key concern is on modeling and evaluating potential losses occurring with extremely low probabilities, i.e. tail risks. For example, the Basel committee on banking supervision suggests regulators to require banks holding adequate capital against the tail risk of bank assets measured by the Value-at-Risk (VaR). The VaR refers to high quantile of the loss distribution with an extremely low tail probability.[1] Estimating such risk measures thus relies on modeling the tail region of distribution functions of asset values. To serve such a purpose, statistical tools stemming from Extreme Value Theory (EVT) are obvious candidates. By investigating data in an intermediate region close to the tail, extreme value statistics employs models to extrapolate intermediate properties to the tail region. Although such an attractive feature of extreme value statistics

[1]In the revised Basel II accord and the subsequent Basel III accord, the VaR measures for risks on both trading and banking books must be calculated at a 99.9% level.

makes it a popular tool for evaluating tail events in many scientific fields such as meteorology and engineering, it has not yet emerged into as a dominating tool in financial risk management. This is potentially due to some crucial critiques on applying EVT to financial data; see, e.g. Diebold et al. [2000]. The critiques are mainly on two issues: the difficulty in selecting the intermediate region in estimation and the validity of the maintained assumptions in EVT for financial data. This paper tries to deal with the two critiques simultaneously and provide adapted EVT methods that overcome the two issues jointly.

We start with explaining the problem on selecting the intermediate region in estimation. Extreme value statistics usually use only observations in an intermediate region. This has been achieved by selecting the highest (or lowest when dealing with lower tail) $k = k(n)$ observations in a sample with size n. The problem on selecting k is sometimes referred to as "selecting the cutoff point". Theoretically, the statistical properties of EVT-based estimators are established for k such that $k \to \infty$ and $k/n \to 0$, as $n \to \infty$. In applications with a finite sample size, it is necessary to investigate how to choose the number of high observations used in estimation. For financial practitioner, two difficulties arise: firstly, there is no straightforward procedure for the selection; secondly, the performance of the EVT estimators is rather sensitive to this choice. More specifically, there is a bias-variance tradeoff: with low level of k, the estimation variance is at a high level which may not be acceptable for application; by increasing k, i.e. using progressively more data, the variance is reduced, but at the cost of an increasing bias.

Recent developments in extreme value statistics provide two types of solutions for selecting the cutoff point. The first type of solutions aims to find the optimal cutoff point that balances the bias and variance assuming that the bias term in the asymptotic distribution is finite; see e.g. Danielsson et al. [2001], Drees and Kaufmann [1998] and Guillou and Hall [2001]. The second type of solutions corrects the bias under allowing that the bias term in the asymptotic distribution is at an infinite level, see e.g. Gomes et al. [2008]. Comparing with the optimal cutoff point method, the bias correction procedure usually requires additional assumptions, such as the third order condition. Nevertheless, it is preferred to the optimal cutoff point approach because of the following relative advantages. First, since bias correction methods allow for an infinite bias term in the asymptotic distribution, they correspondingly allow for choosing a higher level of k than that chosen in the optimal cutoff point approach. Second, by choosing a larger k, bias correction methods result in a lower level of estimation variance with no asymptotically bias. Third, in practice, bias correction procedures lead to estimates that are less sensitive to the choice of k. This mitigates the difficulty in the selection of the cutoff point.

The other criticism on applying extreme value statistics to financial data is on the fact that most existing EVT methods require independent and identically distributed (i.i.d.) observations whereas financial time series exhibits obvious serial dependence feature such as volatility clustering. This issue has been addressed in works dealing with weakly serial dependence, see, e.g. Hsing [1991] and Drees [2000]. The main message from these studies is that usual EVT methods are still valid, only the asymptotic variance of estimators may differ from that in the i.i.d. case.

Although the selection of the cutoff point and the serial dependence in data have been handled separately, the literature addressing these two issues are mutually exclusive. In the bias correction literature, it is always assumed that the observations form an i.i.d. sample; in the literature on dealing with serial dependence, the choice of k is assumed to be sufficiently low such that there is no asymptotic bias. Therefore, it is still an open question whether we can apply the bias correction technique to datasets that exhibit weakly serial dependence. This is what we tend to address in this paper.

We consider bias correction procedure on estimating the extreme value index and high quantiles for $\beta-$mixing stationary time series with common heavy-tailed distribution. The bias term stems from the approximation of the tail region of distribution functions. In EVT, a second order condition is often imposed to characterize such an approximation. Such a condition is almost indispensable for establishing asymptotic properties of estimators. To correct the bias, one needs to estimate the second order scale function, the function A in (3) below. The existing literature is restricted to the case $A(t) = Ct^{\rho}$ with constants $C \neq 0$ and $\rho < 0$. The estimation of C requires extra conditions. Instead we estimate the function A in a non-parametric way which makes the analysis and application smoother.

The asymptotically unbiased estimators we obtain have the following advantages. Firstly, it allows serial dependence in the observations. Secondly, one may apply the unbiased estimator with a higher value of k, which reduces the asymptotic variance and ultimately the estimation error thanks to the bias correction feature. Thirdly, the theoretical range of potential choices of k is larger for our asymptotically unbiased estimators than for the original estimators. This makes the choice of k less crucial. All these features become apparent in simulation and application.

The paper is organized as follows. Under a simplified model without serial dependence, Section 2 presents the bias correction idea for the Hill estimator. Section 3 presents the general model with serial dependence, particularly, the regulatory conditions we are dealing with. Section 4 defines the asymptotically unbiased estimators of the extreme value index and quantiles. In addition, we state the main theorems on the asymptotic normality of these two estimators. The bias correction procedure and the serial dependence structure have a joint impact on the asymptotic variances of the estimators. Section 5 discusses such a joint impact for several examples. Section 6 demonstrates finite sample performance of the asymptotically unbiased estimators based on simulation. An application to estimate the VaR of daily returns on the Dow Jones Industrial Average Index is given in Section 7. All proofs are postponed to Appendix A.

2 The idea of bias correction under independence

For the sake of simplicity, we first introduce our bias correction idea under the assumption of independent and identically distributed (i.i.d.) observations in this section. We will show later that our bias correction procedure also works for $\beta-$mixing series.

2.1 The origin of bias

Let $\{X_1, X_2, \ldots\}$ be an i.i.d. sequence of random variables with a common distribution function F. We assume that this distribution function belongs to the domain of attraction with a positive extreme value index. We present the domain of attraction condition with respect to the quantile function $U := (1/1-F)^{\leftarrow}$, where $^{\leftarrow}$ denotes the left-continuous inverse function. That is, there exists a positive number γ such that

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^{\gamma}, \qquad x > 0\ . \tag{1}$$

Such a distribution function F is also referred as a heavy-tailed distribution. The relation (1) governs how a high quantile, say $U(tx)$, can be extrapolated from an intermediate quantile $U(t)$. Clearly, estimating the extreme value index γ is a major step in estimating high quantiles.

In the heavy-tailed case, Hill [1975] proposes the following estimator of the parameter γ,

$$\hat{\gamma}_k := \frac{1}{k}\sum_{i=1}^{k} \log X_{n-i+1,n} - \log X_{n-k,n}\ , \tag{2}$$

where $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$ are the order statistics and k is an intermediate sequence such that $k \to \infty$ and $k/n \to 0$ as $n \to \infty$.

To obtain the asymptotic normality of the Hill estimator (and most other estimators in EVT), it is necessary to quantify the speed of convergence in (1). We thus assume a second order condition on the function U as follows. Suppose that there exist a positive or negative function A with $\lim_{t\to\infty} A(t) = 0$ and a real number $\rho \leq 0$ such that

$$\lim_{t\to\infty} \frac{\frac{U(tx)}{U(t)} - x^{\gamma}}{A(t)} = x^{\gamma}\frac{x^{\rho}-1}{\rho}\ ,$$

for all $x > 0$. It is equivalent to

$$\lim_{t\to\infty} \frac{\log U(tx) - \log U(t) - \gamma \log x}{A(t)} = \frac{x^{\rho}-1}{\rho}\ ; \tag{3}$$

see, for instance de Haan and Ferreira [2006, Proof of Theorem 3.2.5]. The parameter ρ controls the speed of convergence, both of the sample maximum towards an extreme value distribution and for the extreme value estimators towards a normal distribution. Larger absolute value of ρ means better speed of convergence. This is illustrated in the last paragraph of Section 2.2.

The estimator $\hat{\gamma}_k$ is consistent under the domain of attraction condition (1). Under the second order condition (3), the asymptotic normality can be established for i.i.d. observations as

$$\sqrt{k_\lambda}\left(\hat{\gamma}_{k_\lambda} - \gamma\right) \xrightarrow{d} N\left(\frac{\lambda}{1-\rho}, \gamma^2\right)\ , \tag{4}$$

if the intermediate sequence k_λ satisfies

$$\lim_{n\to\infty} \sqrt{k_\lambda} A(n/k_\lambda) = \lambda, \tag{5}$$

where λ is a finite constant. This condition imposes an upper bound on the speed at which k_λ goes to infinity. The asymptotic bias for the Hill estimator is consequently given by the term $\frac{\lambda}{1-\rho}$.

To obtain an asymptotically unbiased estimator, we will first estimate the bias term and then subtract that from $\hat{\gamma}_k$. The asymptotically unbiased estimator is then given as $\hat{\gamma}_k - \widehat{\text{Bias}_k}$, where

$$\text{Bias}_k := \frac{A(n/k)}{1-\rho} . \tag{6}$$

A formal definition of the asymptotically unbiased estimator is given in Equation (12) below. It is clear that the second order parameter ρ plays an important role in the bias term.

2.2 Estimating the bias term

The estimation of the bias term requires estimating the second order parameter ρ and the second order scale function, $A(n/k)$, appearing in the condition (3). The parameter ρ controls the speed of convergence of most γ estimators. In the following we restrict the study to the case $\rho < 0$ because the estimation of the bias term exploits the regular variation feature of the A function whereas the case of slowly variation ($\rho = 0$) is difficult to be handled. In the literature of bias correction, in order to establish the asymptotic property of estimators of ρ, it is necessary to choose a higher intermediate sequence $k_\rho = k_\rho(n)$ such that $k_\rho \to \infty$, $k_\rho/n \to 0$ and

$$\sqrt{k_\rho}\, A(n/k_\rho) \to \infty , \tag{7}$$

as $n \to \infty$, see e.g. Gomes et al. [2002]. This provides a lower bound to the speed at which k_ρ goes to infinity. Also, a third order condition is useful. Suppose that there exist a positive or negative function B with $\lim_{t\to\infty} B(t) = 0$ and a real number $\rho' \le 0$ such that

$$\lim_{t\to\infty} \frac{1}{B(t)} \left\{ \frac{\log U(tx) - \log U(t) - \gamma \log x}{A(t)} - \frac{x^\rho - 1}{\rho} \right\} = \frac{1}{\rho'} \left\{ \frac{x^{\rho+\rho'} - 1}{\rho + \rho'} - \frac{x^\rho - 1}{\rho} \right\} , \tag{8}$$

for all $x > 0$. If the observations are i.i.d., the asymptotic normality of all existing estimators of ρ, including that of the one we use in (11) below, holds under the condition (8) and with a sequence k_ρ such that as $n \to \infty$, $k_\rho \to \infty$, $k_\rho/n \to 0$ and

$$\sqrt{k_\rho} A(n/k_\rho) \to \infty, \sqrt{k_\rho} A^2(n/k_\rho) \to \lambda_1, \sqrt{k_\rho} A(n/k_\rho) B(n/k_\rho) \to \lambda_2 , \tag{9}$$

where λ_1 and λ_2 are both finite constants; see for instance Gomes et al. [2002] and Ciuperca and Mercadier [2010]. Here, since we are going to deal with $\beta-$mixing series, we need to re-establish the asymptotic property of the ρ estimator. The details are left to Appendix A.

In order to avoid extra bias stemming from the third order condition, the k sequence we use for the asymptotically unbiased estimator of the extreme value index is of a lower order, compared to the k_ρ sequence. More specifically, we use a sequence k_n such that as $n \to \infty$, $k_n \to \infty$, $k_n/k_\rho \to 0$ and

$$\sqrt{k_n}A(n/k_n) \to \infty, \sqrt{k_n}A^2(n/k_n) \to 0, \sqrt{k_n}A(n/k_n)B(n/k_n) \to 0 \, . \tag{10}$$

Comparing our asymptotically unbiased estimator with the original Hill estimator, the k sequences used for estimation are at different level. The conditions on k_n and k_λ imply that $k_n/k_\lambda \to +\infty$ as $n \to \infty$. Since the asymptotic variance of both the asymptotically unbiased estimator and the original Hill estimator is of an order $1/k$, using a sequence k_n increasing faster than k_λ leads to a lower asymptotic variance of our asymptotic unbiased estimator compared to that of the original Hill estimator.

In addition, the k sequence used for the asymptotically unbiased estimator is more flexible in the following sense. The third order condition (8) implies that A and B are regularly varying functions with index ρ and ρ' respectively. Consider the special case that $A(t) \sim Ct^\rho$ and $B(t) \sim Dt^{\rho'}$ as $t \to \infty$ for some constant C and D. Then the condition that $\sqrt{k_\lambda}A(n/k_\lambda) \to \lambda$ restricts the level of k_λ as $k_\lambda = O\left(n^{\frac{2\rho}{2\rho-1}}\right)$, whereas condition (10) implies that $k_n = O\left(n^\tau\right)$ for any $\tau \in \left(\frac{2\rho}{2\rho-1}, \frac{2(\rho+\max(\rho,\rho'))}{2(\rho+\max(\rho,\rho'))-1}\right)$.

3 The serial dependence conditions

In this section, we present the serial dependence conditions on the time series we are going to deal with. The serial dependence structure follows from the so-called $\beta-$mixing conditions. The $\beta-$mixing conditions have been introduced by Rootzén [1995], Drees [2000, 2003] and Rootzén [2009] as follows. Let $\{X_1, X_2, \ldots\}$ be a stationary time series with common distribution function F. Let $\mathcal{B}_i^j$ denote the σ-algebra generated by $X_i, \ldots, X_j$. The sequence is said to be β-mixing or absolutely regular if

$$\beta(m) := \sup_{\ell \geq 1} \mathbb{E}\left(\sup_{E \in \mathcal{B}_{\ell+m+1}^\infty} \left|\mathbb{P}(E|\mathcal{B}_1^\ell) - \mathbb{P}(E)\right|\right) \to 0$$

as $m \to \infty$. The constants $\beta(m)$ are called the β-mixing constants of the sequence.

The asymptotic normality of the original Hill estimator has been established for β-mixing sequences in Drees [2000, 2003] with some mild extra conditions. With a sequence k_λ such that $\sqrt{k_\lambda}A(n/k_\lambda) \to \lambda$ as $n \to \infty$, it is proved that

$$\sqrt{k_\lambda}\left(\hat{\gamma}_{k_\lambda} - \gamma\right) \xrightarrow{d} N\left(\frac{\lambda}{1-\rho}, \sigma^2\right) \, ,$$

where σ^2 is equal to γ^2 under independence but is more complicated otherwise. The extra conditions for establishing the asymptotic normality of the Hill estimator are the following list of *regulatory* *conditions.* Suppose there exist a constant $\varepsilon > 0$, a function $r(\cdot,\cdot)$ and a sequence $\ell = \ell_n$ such that as $n \to \infty$,

(a) $\frac{\beta(\ell)}{\ell} n + \ell k^{-1/2} \log^2 k \to 0$,

(b) $\frac{n}{\ell k} \mathrm{Cov}\left(\sum_{i=1}^{\ell} \mathbf{1}_{\{X_i > F^{-1}(1-kx/n)\}}, \sum_{i=1}^{\ell} \mathbf{1}_{\{X_i > F^{-1}(1-ky/n)\}}\right) \to r(x,y)$, for any $0 \le x, y \le 1+\varepsilon$,

(c) for some constant C,

$$\frac{n}{\ell k} \mathbb{E}\left[\left(\sum_{i=1}^{\ell} \mathbf{1}_{\{F^{-1}(1-ky/n) < X_i \le F^{-1}(1-kx/n)\}}\right)^4\right] \le C(y-x) ,$$

for any $0 \le x < y \le 1+\varepsilon$ and $n \in \mathbb{N}$.

Drees [2000] shows that the condition (a) is fulfilled if the original time series $\{X_1, X_2, \cdots\}$ is geometrically $\beta-$mixing, i.e. $\beta(m) = O(\eta^m)$ for some $\eta \in (0,1)$. In that case, one may take $\ell_n = [-2\log n / \log \eta]$. Drees [2003] remarks that the condition (b) holds if all vectors (X_1, X_{1+m}) belong to the domain of attraction of a bivariate extreme value distribution. In that case, for any sequence k, one may take a sequence ℓ such that $\ell k/n \to 0$ as $n \to \infty$. The limit function $r(x,y)$ depends only on the tail dependence structure of (X_1, X_{1+m}) for $m \in \mathbb{N}$. These two sufficient version of conditions (a) and (b) hold for some known time series model, namely the ARMA, ARCH and GARCH models, see the examples in Section 5 below. Lastly, the condition (c) has been verified for these time series models as well. In addition, for all these models, it is only necessary to have $k = o(n^{\zeta})$ for some $\zeta < 1$ as $n \to \infty$ in order to satisfy the regulatory conditions. This is compatible with the requirement on the sequence k_λ in extreme value analysis as follows. Under the second order condition, $|A(t)|$ is regularly varying with index ρ. Therefore, given any $\varepsilon > 0$, for sufficiently large t, we have that $|A(t)| > Ct^{\rho-\varepsilon}$ for some positive constant C; see inequality (B.1.19) in de Haan and Ferreira [2006]. Together with the condition (5), we get that $k_\lambda = o(n^{\zeta})$ for any $\zeta > \frac{2\rho-\varepsilon}{2\rho-1-\varepsilon}$. Therefore, the sequence k_λ is compatible with condition (c).

We intend to correct the bias while allowing the observations to follow the $\beta-$mixing condition and the regulatory conditions. Since the asymptotic bias of the original Hill estimator under serial dependence has the same form as in (6), we can construct an asymptotically unbiased estimator for β-mixing sequences with exactly the same form as in the independence case. Nevertheless, due to the serial dependence, the asymptotic property of the estimator has to be reestablished. This is what we do in the next section.

4 Main Results

We start by introducing the estimator of the second order parameter. Then we state our main results on the asymptotic properties of the asymptotic unbiased estimators of the extreme value index and high quantiles.

4.1 Estimating the second order parameter

We adopt the notations of Gomes et al. [2002] as follows. For any positive number α, denote

$$M_k^{(\alpha)} := \frac{1}{k}\sum_{i=1}^{k}\left(\log X_{n-i+1,n} - \log X_{n-k,n}\right)^{\alpha} \ ,$$

$$R_k^{(\alpha)} := \frac{M_k^{(\alpha)} - \Gamma(\alpha+1)\left(M_k^{(1)}\right)^{\alpha}}{M_k^{(2)} - 2\left(M_k^{(1)}\right)^{2}} \ ,$$

$$S_k^{(\alpha)} := \frac{\alpha(\alpha+1)^2\Gamma^2(\alpha)}{4\Gamma(2\alpha)}\frac{R_k^{(2\alpha)}}{\left(R_k^{(\alpha+1)}\right)^{2}} \ ,$$

$$s^{(\alpha)}(\rho) := \frac{\rho^2\left(1-(1-\rho)^{2\alpha} - 2\alpha\rho(1-\rho)^{2\alpha-1}\right)}{\left(1-(1-\rho)^{\alpha+1} - (\alpha+1)\rho(1-\rho)^{\alpha}\right)^{2}} \ .$$

Then the estimator of the second order parameter ρ is defined as

$$\hat{\rho}_k^{(\alpha)} := \left(s^{(\alpha)}\right)^{\leftarrow}(S_k^{(\alpha)}) \ . \tag{11}$$

4.2 Asymptotically unbiased estimator of the extreme value index

We now write explicitly the asymptotically unbiased estimator of the extreme value index. Let k_ρ and k_n, satisfying (9) and (10), be the number of observations selected for estimating ρ and γ respectively. For some positive real number α, we define the asymptotically unbiased estimator as

$$\hat{\gamma}_{k_n,k_\rho,\alpha} := \hat{\gamma}_{k_n} - \frac{M_{k_n}^{(2)} - 2\hat{\gamma}_{k_n}^2}{2\hat{\gamma}_{k_n}\hat{\rho}_{k_\rho}^{(\alpha)}(1-\hat{\rho}_{k_\rho}^{(\alpha)})^{-1}} \ , \tag{12}$$

where $\hat{\gamma}_{k_n}$ denotes the original Hill estimator as in (2).

The following theorem shows the asymptotic normality of our asymptotically unbiased estimator for $\beta-$mixing series. The consistency of the estimator could be obtained under the second order condition without requiring the third order condition.

Theorem 4.1. *Suppose that $\{X_1, X_2, \ldots\}$ is a stationary β-mixing time series with continuous common marginal distribution function F. Assume that F satisfies the third order condition (8) with parameters $\gamma > 0$, $\rho < 0$ and $\rho' \leq 0$. Suppose that the two intermediate sequences k_ρ and k_n satisfy the conditions in (9) and (10) respectively. Suppose that the regulatory conditions (a)-(c) hold with the intermediate sequence k_n. Then,*

$$\sqrt{k_n}\left(\hat{\gamma}_{k_n,k_\rho,\alpha} - \gamma\right) \xrightarrow{d} N\left(0,\sigma^2\right) \ ,$$

where

$$\sigma^2 := \frac{\gamma^2}{\rho^2}\left((2-\rho)^2 c_{1,1} + (1-\rho)^2 c_{2,2} + 2(2-\rho)(\rho-1)c_{1,2}\right) ,$$

with

$$c_{i,j} := \iint_{[0,1]^2} (-\log s)^{i-1}(-\log t)^{j-1}\left\{\frac{r(s,t)}{st} - \frac{r(s,1)}{s} - \frac{r(1,t)}{t} + r(1,1)\right\} ds\, dt ,$$

and $r(\cdot,\cdot)$ defined in the regulatory condition (b).

Compared to the original Hill estimator, we use different k sequences, namely k_n and k_ρ, in the asymptotically unbiased estimator $\hat{\gamma}_{k_n,k_\rho,\alpha}$. These k sequences are compatible with the regulatory conditions. Recall that the third order condition (8) implies that A^2 and AB are regularly varying functions with index 2ρ and $\rho+\rho'$ respectively. Conditions (9) and (10) ensures that $k_n, k_\rho = o(n^\zeta)$ for some $\zeta < 1$ and consequently the compatibility of these two sequences with the regulatory conditions. In general, as long as the original k_λ sequence is compatible with the regulatory conditions, so are k_n and k_ρ.

We remark that our estimator is also valid as an asymptotically unbiased estimator of the extreme value index when the observations are i.i.d.. In that case, the result is simplified to

$$\sqrt{k_n}\left(\hat{\gamma}_{k_n,k_\rho,\alpha} - \gamma\right) \xrightarrow{d} N\left(0, \frac{\gamma^2}{\rho^2}\left\{\rho^2 + (1-\rho)^2\right\}\right) .$$

4.3 Asymptotically unbiased estimator of high quantiles

We consider the estimation of high quantiles. High quantile refers to the quantile at a probability level $(1-p)$, where the tail probability $p = p_n$ depends on the sample size n: as $n \to \infty$, $p_n = O(1/n)$. The goal is to estimate the quantile $x(p) = U(1/p)$. In extreme case such that $np_n < 1$, it is not possible to have non-parametric estimate of such a quantile.

We derive the following asymptotically unbiased estimator:

$$\hat{x}_{k_n,k_\rho,\alpha}(p) := X_{n-k_n,n}\left(\frac{k_n}{np}\right)^{\hat{\gamma}_{k_n,k_\rho,\alpha}}\left(1 - \frac{(M_{k_n}^{(2)} - 2\hat{\gamma}_{k_n}^2)(1-\hat{\rho}_{k_\rho}^{(\alpha)})^2}{2\hat{\gamma}_{k_n}\{\hat{\rho}_{k_\rho}^{(\alpha)}\}^2}\right) .$$

The first factor $X_{n-k_n,n}\left(\frac{k_n}{np}\right)^{\hat{\gamma}_{k_n,k_\rho,\alpha}}$ in this formula follows a similar structure as the quantile estimator in Weissman [1978]. Having the additional term is to correct the extra bias induced by using a high level k_n; See a similar treatment in Cai et al. [2013] for the quantile estimator using an asymptotically unbiased probability-weighted-moment approach.

The following theorem shows the asymptotic normality of the quantile estimator $\hat{x}_{k_n,k_\rho,\alpha}(p)$.

Theorem 4.2. *Suppose that $\{X_1, X_2, \dots\}$ is a stationary β-mixing time series with continuous common marginal distribution function F. Assume that F satisfies the third order condition (8) with parameters $\gamma > 0$, $\rho < 0$ and $\rho' \leq 0$. Suppose that the two intermediate sequences k_ρ and k_n*

satisfy the conditions in (9) *and* (10) *respectively. Assume in addition that* $n \to \infty$, $np_n/k_n \to 0$ *and* $\log(np_n)/\sqrt{k_n} \to 0$. *Suppose that the regulatory conditions (a)-(c) hold with* k_n. *Then*

$$\frac{\sqrt{k_n}}{\log(k_n/(np_n))}\left(\frac{\hat{x}_{k_n,k_\rho,\alpha}(p_n)}{x(p_n)} - 1\right) \xrightarrow{d} N\left(0, \sigma^2\right) ,$$

with σ^2 *as defined in Theorem 4.1.*

5 Examples

In our framework, we model the serial dependence by the $\beta-$mixing condition and the extra regulatory conditions. In this section, we give a few examples that satisfy those conditions. The referred studies below have documented that these examples satisfy the regulatory conditions for any sequence k such that $k = o(n^\zeta)$ for some $\zeta < 1$ as $n \to \infty$.

- The k-dependent process and the autoregressive (AR) process, AR(1): Rootzén [1995], Drees [2003], Rootzén [2009];
- The AR(p) processes and the infinite moving averages (MA) processes: Resnick and Stărică [1997], Drees [2002];
- The finite MA processes: Hsing [1991], Rootzén [1995], Drees [2002], Rootzén [2009];
- The autoregressive conditional heteroskedasticity process, ARCH(1): Drees [2002, 2003];
- The generalized autoregressive conditional heteroskedasticity (GARCH) processes: Stărică [1999], Drees [2000].

We review some simple cases of these processes and provide the comparison of the asymptotic variances under dependence to that under independence, and to that of the original Hill estimator under serial dependence.

5.1 Autoregressive model

Consider first the stationary solution of the following AR(1) equation

$$X_i = \theta X_{i-1} + Z_i , \tag{13}$$

for some $\theta \in (0, 1)$ and i.i.d. random variables Z_i. The distribution function of the innovation is denoted by F_Z. Assume that F_Z admits a positive Lebesgue density that satisfies the Lipschitz condition of order 1 (Billingsley [1979], pp. 418). Suppose that as $x \to \infty$,

$$1 - F_Z(x) \sim px^{-1/\gamma}\ell(x) \text{ and } F_Z(-x) \sim qx^{-1/\gamma}\ell(x) \tag{14}$$

for some slowly varying function ℓ and $p = 1 - q \in (0, 1)$. Then from Section 3.2 of Drees [2003] we get that $1 - F(x) \sim d_\theta\,(1 - F_Z(x))$ as $x \to \infty$, where $d_\theta = (1 - \theta^{1/\gamma})^{-1}$. Furthermore, the regulatory conditions hold with

$$r(x, y) = x \wedge y + \sum_{m=1}^{\infty} \{c_m(x, y) + c_m(y, x)\},$$

where $c_m(x, y) = x \wedge y\theta^{m/\gamma}$.

Let us denote by $\sigma^2(\theta, \gamma, \rho)$ the asymptotic variance of $\sqrt{k}(\hat{\gamma}_{k,k_\rho,\alpha} - \gamma)$. First, we compare the asymptotic variance under model (13) with that under independence by calculating the ratio $\sigma^2(\theta, \gamma, \rho)/\sigma^2(0, \gamma, \rho)$. Second, we compare $\sigma^2(\theta, \gamma, \rho)$ with the asymptotic variance of the original Hill estimator under serial dependence, σ_H^2, when using the same k sequence. From Drees [2000], we get that under serial dependence $\sqrt{k}(\hat{\gamma}_H - \gamma)$ converges to a normal distribution with asymptotic variance $\sigma_H^2 = \gamma^2 r(1, 1)$. The two ratios are given as follows.

$$\begin{aligned}
\frac{\sigma^2(\theta, \gamma, \rho)}{\sigma^2(0, \gamma, \rho)} &= 1 + \frac{2\theta^{1/\gamma}}{1 - \theta^{1/\gamma}} + \frac{2\rho(1-\rho)}{1 - 2\rho(1-\rho)} \frac{\theta^{1/\gamma} \log \theta^{1/\gamma}}{(1 - \theta^{1/\gamma})^2}\ , \\
\frac{\sigma^2(\theta, \gamma, \rho)}{\sigma_H^2} &= \frac{1}{\rho^2}\left(1 - 2\rho(1-\rho) + 2\rho(1-\rho)\frac{\theta^{1/\gamma} \log \theta^{1/\gamma}}{(1 - \theta^{1/\gamma})^2 + 2\theta^{1/\gamma}(1 - \theta^{1/\gamma})}\right)\ .
\end{aligned}$$

In the first row of Figure 1, we plot these ratios against the extreme value index γ for different values of the parameters θ and ρ. From Figure 1a, we note that the variation of the first ratio is mainly due to that of θ. The parameter ρ plays a relative minor role. We further give a numerical illustration with $\gamma = 1$ and $\rho = -1$. With i.i.d. observations, the asymptotic variance of $\sqrt{k}(\hat{\gamma}_{k,k_\rho,\alpha} - \gamma)$ is 5. Instead, if the observations follow the AR(1) model with $\theta = 0.5$, then the asymptotic variance of $\sqrt{k}(\hat{\gamma}_{k,k_\rho,\alpha} - \gamma)$ is close to 20. Hence, overlooking the serial dependence may severely underestimate the range of confidence intervals.

Differently, we observe from Figure 1b that the variation of the second ratio is mainly due to that of ρ. Although this ratio is greater than one, it does not imply that the asymptotically unbiased estimator has a higher asymptotic variance, because the current comparison is conducted using the same k level for both estimators, whereas the k value used in the asymptotically unbiased estimator can be at a much higher level than that used for the Hill estimator. Theoretically the conditions on k_n and k_λ guarantees that $k_n/k_\lambda \to +\infty$. Thus the variance of our estimator is at a lower level asymptotically. Practically, if we consider the example that $\rho = -1$, then the ratio is in between 5 and 7. Under such an example, if we use a k_n in the asymptotically unbiased estimator seven times higher than k_λ used for the original Hill estimator, we will get an estimator with lower variance. If the level of ρ is more close to zero, then the ratio will be at a higher level. Correspondingly, one needs a higher level of k_n to offset the higher ratio. Nevertheless, together with the fact that the asymptotically unbiased estimator does not suffer from the bias issue, it may still perform better in terms of having a lower root mean squared error. Such a feature will show up in the simulation studies in Section 6 below.

5.2 Moving average model

Consider now the stationary solution of the MA(1) equation

$$X_i = \theta Z_{i-1} + Z_i \ , \tag{15}$$

where the innovation Z satisfies the same conditions as in the AR(1) model in the previous subsection. Again from Section 3.2 of Drees [2003] we get that $1 - F(x) \sim d_\theta\,(1 - F_Z(x))$ as $x \to \infty$, where $d_\theta = 1 + \theta^{1/\gamma}$. One can also compute

$$r(x, y) = x \wedge y + (1 + \theta^{1/\gamma})^{-1}(x \wedge y\theta^{1/\gamma} + y \wedge x\theta^{1/\gamma}) \ .$$

We calculate the two ratios when comparing the asymptotic variance of the asymptotically unbiased estimator under serial dependence to that under independence, and that of the original Hill estimator under dependence as follows.

$$\begin{aligned}
\frac{\sigma^2(\theta,\gamma,\rho)}{\sigma^2(0,\gamma,\rho)} &= 1 + \frac{2\theta^{1/\gamma}}{1+\theta^{1/\gamma}} + \frac{2\rho(1-\rho)}{1-2\rho(1-\rho)}\frac{\theta^{1/\gamma}\log\theta^{1/\gamma}}{1+\theta^{1/\gamma}} \ , \\
\frac{\sigma^2(\theta,\gamma,\rho)}{\sigma_H^2} &= \frac{1}{\rho^2}\left(1 - 2\rho(1-\rho) + 2\rho(1-\rho)\frac{\theta^{1/\gamma}\log\theta^{1/\gamma}}{(1+\theta^{1/\gamma}) + 2\theta^{1/\gamma}}\right) \ .
\end{aligned}$$

In the second row of Figure 1, we plot the variations of these ratios with respect to the extreme value index γ for different values of the parameters θ and ρ. The general feature is comparable to that observed from the first row. A notable difference between Figures 1a and 1c is that although the ratios are both increasing in θ and the absolute value of ρ, their convexities with respect to γ are different in the two models: we observe a concave (resp. convex) relation in γ under the MA(1) (resp. AR(1)) model.

5.3 Generalized autoregressive conditional heteroskedasticity model

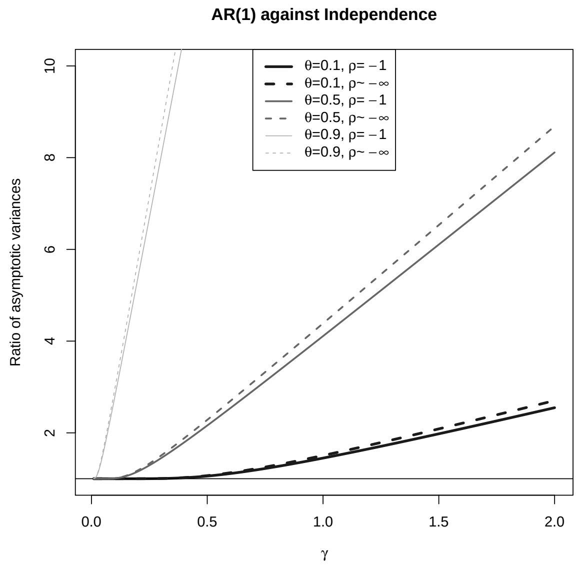
Consider the stationary solution to the following recursive system of equations

$$\begin{cases} X_t &= \varepsilon_t \sigma_t, \\ \sigma_t^2 &= \lambda_0 + \lambda_1 X_{t-1}^2 + \lambda_2 \sigma_{t-1}^2, \end{cases}$$

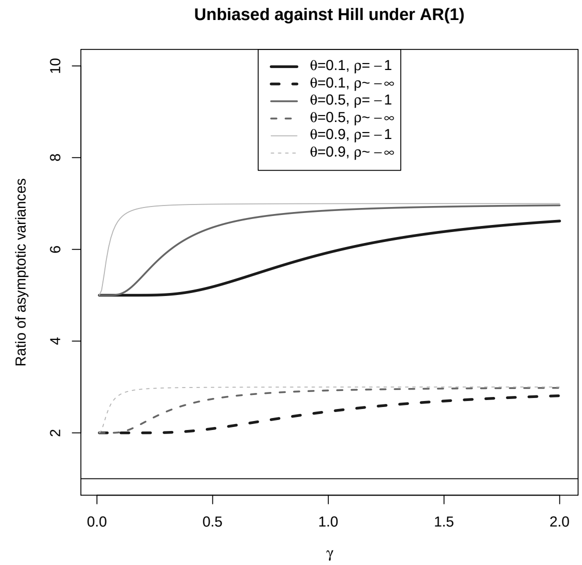
where ε_t are i.i.d. innovations with zero mean and unit variance. The stationary solution of this GARCH(1,1) model, X_t, follows a heavy-tailed distribution, even if the innovations ε_t are normally distributed, see Kesten [1973] and Goldie [1991]. The extreme value index of the GARCH(1,1) model can be derived from the Kesten theorem on stochastic difference equations, see Kesten [1973]. Nevertheless, the calculation is not explicit.

In addition, the stationary GARCH(1,1) series satisfies the $\beta-$mixing condition and the regulatory conditions, see Stărică [1999] and Drees [2000]. Thus, it can be considered as an example

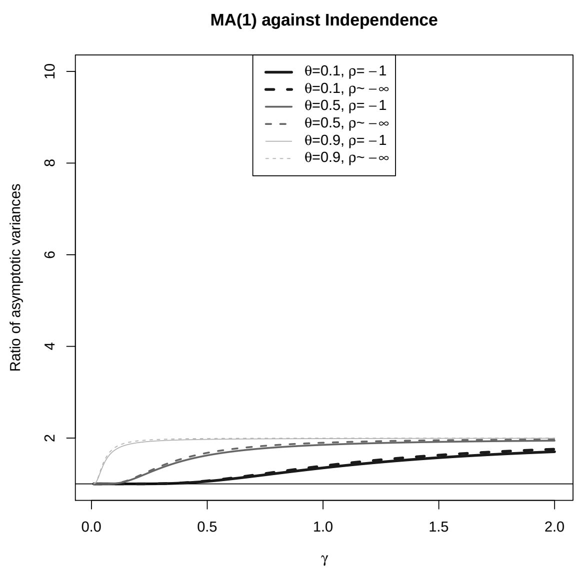
Figure 1: **Ratios between asymptotic variances.** Figure 1a shows the ratio between the asymptotic variance of the asymptotically unbiased estimator under the AR(1) model and that under the i.i.d. case. Figure 1b shows the ratio between the asymptotic variance of the asymptotically unbiased estimator and that of the original Hill estimator, under the AR(1) model. Figure 1c and 1d show the corresponding ratios when the serial dependence is modeled by the MA(1) model.



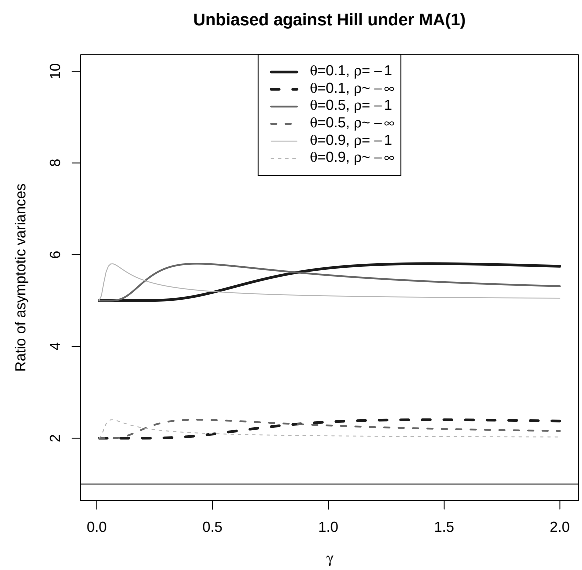
(a) Comparison with the i.i.d. case: AR(1)



(b) Comparison with the Hill estimator: AR(1)



(c) Comparison with the i.i.d. case: MA(1)



(d) Comparison with the Hill estimator: MA(1) for which we can apply the asymptotically unbiased estimators. Since it is difficult to explicitly calculate the $r(\cdot,\cdot)$ function and consequently the asymptotic variance, we opt to use simulations to show the performance of the asymptotically unbiased estimator under the GARCH model.

6 Simulation

6.1 Data generating processes

The simulations are set up as follows. We consider four data generating processes for simulating the observations used in our simulation study. Suppose Z follows the distribution F_Z given by

$$F_Z(x) = \begin{cases} (1-p)(1-\tilde{F}(-x)) & \text{if } x < 0, \\ 1-p+p\tilde{F}(x) & \text{if } x > 0, \end{cases}$$

where $\tilde{F}$ is the standard Fréchet distribution function: $\tilde{F}(x) = \exp(-1/x)$ for $x > 0$, and $p = 0.75$. Then F_Z belongs to the domain of attraction with extreme value index 1. We construct three time series models based on i.i.d. observations Z_t as follows:

(Model 1) Independence: $X_t = Z_t$ (can be regarded as MA(1) with $\theta = 0$),

(Model 2) AR(1): X_t given by (13) with $\theta = 0.3$,

(Model 3) MA(1): X_t given by (15) with $\theta = 0.3$.

In all three models, the theoretical value of γ is 1. In addition, we construct a GARCH(1,1) model as in Subsection 5.3. We remark that the heavy-tailed feature of the GARCH(1,1) model does not depend on whether the innovations follow a heavy-tailed distribution. Nevertheless, empirical evidence supports using heavy-tailed innovations for modeling financial time series, see, e.g. McNeil and Frey [2000] and Sun and Zhou [2013]. Correspondingly, we use the Student-t distribution as the distribution of innovations.[2] All parameters in the simulated GARCH(1,1) model are equal to the estimates from the real data application in Section 7 .

(Model 4) GARCH(1,1): X_t given as in Subsection 5.3 with $\lambda_0 = 8.26 \times 10^{-7}$, $\lambda_1 = 0.052$, $\lambda_2 = 0.941$. The innovation term follows the standardized Student-t distribution with degree of freedom $\nu = 5.64$.

Following Kesten [1973] we calculate the extreme value index γ of the series in Model 4 at 0.258.

In our simulation study, we also compare the performance of our asymptotically unbiased quantile estimator to that of the original Weissman estimator. For that purpose, we estimate $x(0.001)$ for simulated samples from the four data generating processes. We conduct pre-simulations to get the

[2] In order to get a unit variance, we simply normalize the standard Student-t distribution with degree of freedom ν by its standard deviation $\sqrt{\nu/(\nu-2)}$.

theoretical values of $x(0.001)$: for each model, we simulate 500 samples with sample size 10^6 and obtain 500 estimations of $x(0.001)$. Table 1 reports the median of these 500 values for each model.

Table 1: **Simulated theoretical values of $x(0.001)$ under Model 1–4.**

Model 1	Model 2	Model 3	Model 4
749.80	1072.26	972.85	0.0592

6.2 Estimation procedure

For each data generating process, we simulate $N = 1000$ samples with sample size $n = 1000$ each. Firstly, we focus on the extreme value index γ. We apply both the original Hill estimator and the asymptotically unbiased estimator in (12) to each sample. To apply the asymptotically unbiased estimator we use the following procedure.

- Estimate the second order index ρ by (11) with $\alpha = 2$
  - Denote m as the number of positive observations in the sample. For each k satisfying $k \le \min\left(m-1, \frac{2m}{\log\log m}\right)$, calculate the statistic
  $$S_k^{(2)} = \frac{3}{4}\frac{(M_k^{(4)} - 24\{M_k^{(1)}\}^4)(M_k^{(2)} - 2\{M_k^{(1)}\}^2)}{M_k^{(3)} - 6\{M_k^{(1)}\}^3} .$$
  - If $S_k^{(2)} \in [2/3, 3/4]$, then let
  $$\hat{\rho}_k = \frac{-4 + 6S_k^{(2)} + \sqrt{3S_k^{(2)} - 2}}{4S_k^{(2)} - 3} .$$
  - If $S_k^{(2)} < 2/3$ or $S_k^{(2)} > 3/4$, then $\hat{\rho}_k$ does not exist.
  - The parameter ρ is estimated as $\hat{\rho}_{k_\rho}$ with
  $$k_\rho = \sup\left\{k : k \le \min\left(m-1, \frac{2m}{\log\log m}\right) \text{ and } \hat{\rho}_k \text{ exists}\right\}.$$
- Estimate the extreme value index by (12) for various values of k_n,[3] using $\hat{\rho}_{k_\rho}$.

Here the choice of k_ρ in the first step follows the recommendation in Gomes et al. [2002].

Next, we estimate the high quantile $x(0.001)$ by both the original Weissman estimator and the asymptotically unbiased estimator as in Section 4.3. When applying the asymptotically unbiased estimator for high quantiles, we use the same $\hat{\rho}_{k_\rho}$ as estimated above.

[3] For Model 1–3, we use $k_n = 10, 11, \ldots, 700$, while for Model 4, we use $k_n = 10, 11, \ldots, 450$ due to lower number of positive observations.

After obtaining the estimates in the $N = 1000$ samples as $\hat{\gamma}_k^{(j)}$ for $j = 1, 2, \cdots, N$, we calculate the average absolute bias (ABias) and the root mean square error (RMSE) for the two extreme value index estimators by

$$\text{ABias}_k = \left| \frac{1}{N} \sum_{j=1}^{N} \frac{\hat{\gamma}_k^{(j)}}{\gamma} - 1 \right|, \text{ and } \text{RMSE}_k = \sqrt{\frac{1}{N} \sum_{j=1}^{N} \left(\frac{\hat{\gamma}_k^{(j)}}{\gamma} - 1 \right)^2}.$$

Then we plot the results against the corresponding k values in Figure 2–5 for each model respectively.

Similarly, we obtain the ABias and RMSE for the two high quantile estimators. The ABias and RMSE for the four models are plotted in Figure 6–9 respectively.

6.3 Results

Regarding the estimation of the extreme value index, we observe that even with a rather high level of k, our asymptotically unbiased estimator does not suffer from a significant bias, at least for the first three models; see Figure 2–4. In Model 4, the bias term increases with respect to k, but still stays at a lower level than that of the original Hill estimator; see Figure 5. In addition, we compare the reduction of RMSE when switching from the original Hill estimator to the asymptotically unbiased estimator. Across the first three models, the best levels of RMSE are reached for the largest values of k. In Model 4, the RMSE has a different pattern as k increases. However, the reduction is the most significant in Model 4. Although the lowest achieved RMSE for the asymptotically unbiased estimator is at a comparable level as the lowest RMSE for the original Hill estimator for Model 2 and 3, the decreasing of RMSE with respect to k demonstrated by the asymptotically unbiased estimator allows for a more flexible choice of k compared to the U-shaped RMSE demonstrated by the original Hill estimator.

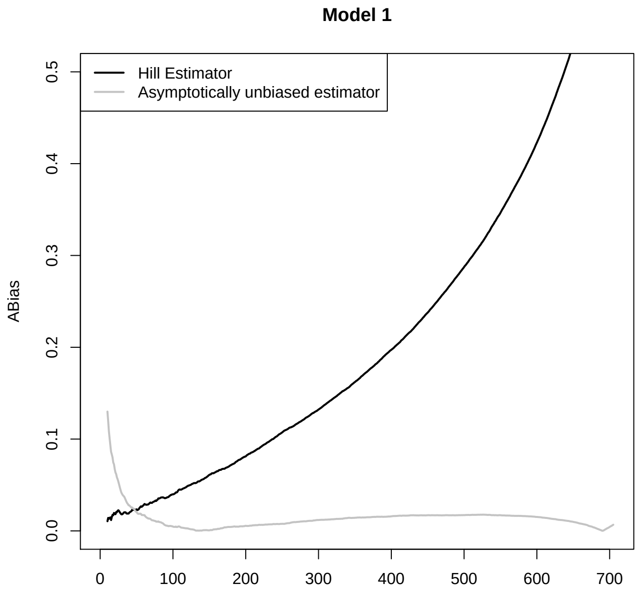
Regarding the estimation of high quantile, we observe from Figure 6–9 that our goal in reducing the bias is well illustrated on finite sample when using large k values. In addition, the RMSE of our asymptotically unbiased quantile estimator stays at a lower level than that of the original Weissman estimator for high levels of k. It is remarkable that the reduction in RMSE is higher for dependent series than for independent series.

To conclude, the simulation studies show that under bias correction, the estimators for extreme value index and high quantile remain stable for a wider range of k values even if the dataset exhibit serial dependence. The bias correction method under serial dependence thus helps to tackle the two major critiques for applying extreme value statistics to financial time series.

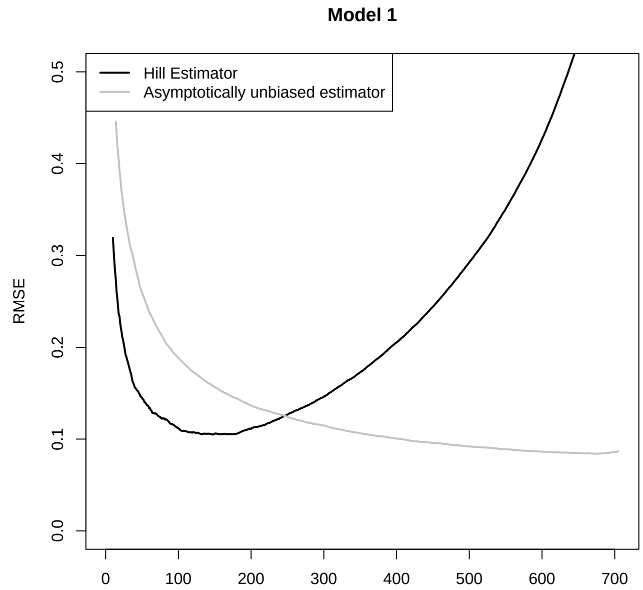
7 Application

We apply the asymptotically unbiased estimators on the extreme value index and high quantiles to evaluate the downside tail risk in the Dow Jones Industrial Average (DJIA) index. We collect the

Figure 2: **Estimating the extreme value index: Model 1.**

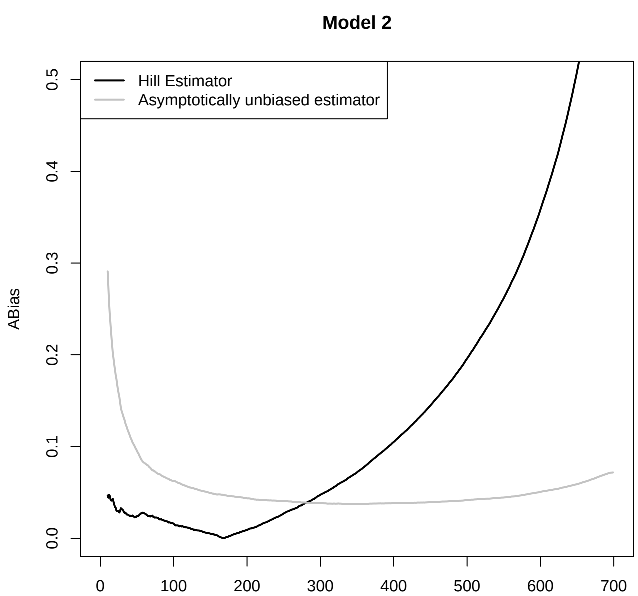


(a) ABias under Model 1.

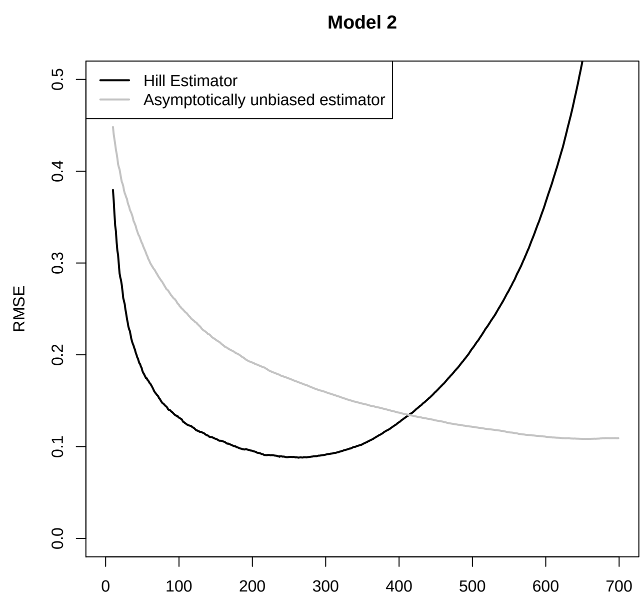


(b) RMSE under Model 1.

Figure 3: **Estimating the extreme value index: Model 2.**

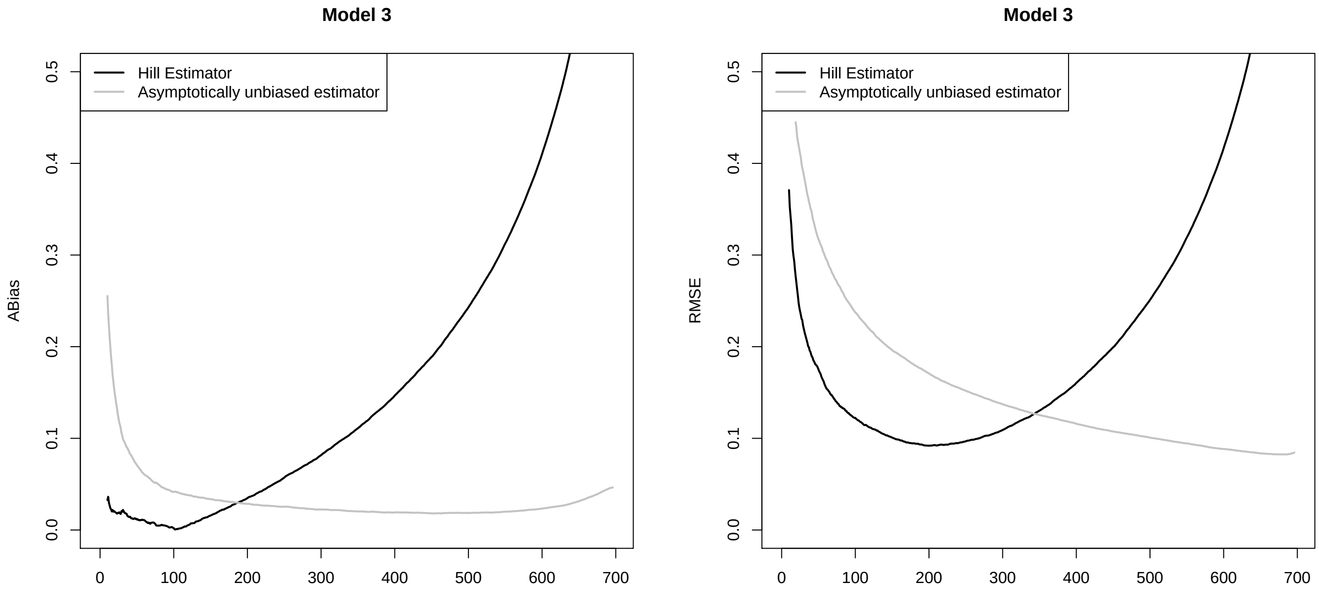


(a) ABias under Model 2.



(b) RMSE under Model 2.

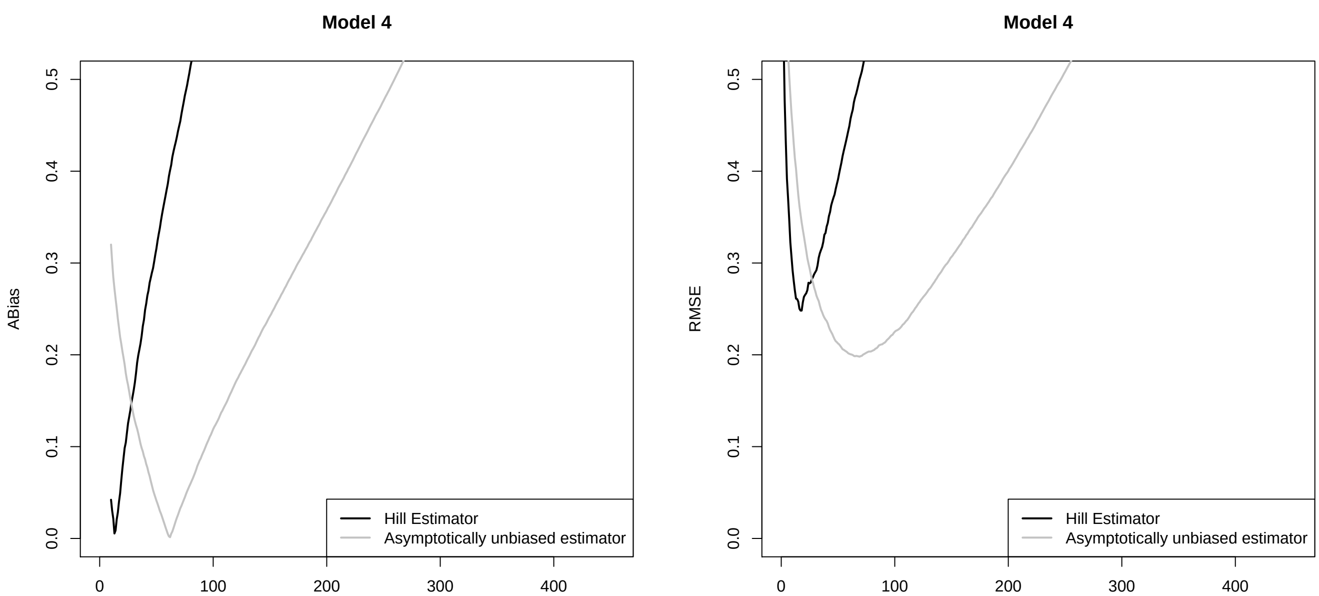
Figure 4: **Estimating the extreme value index: Model 3.**



(a) ABias under Model 3.

(b) RMSE under Model 3.

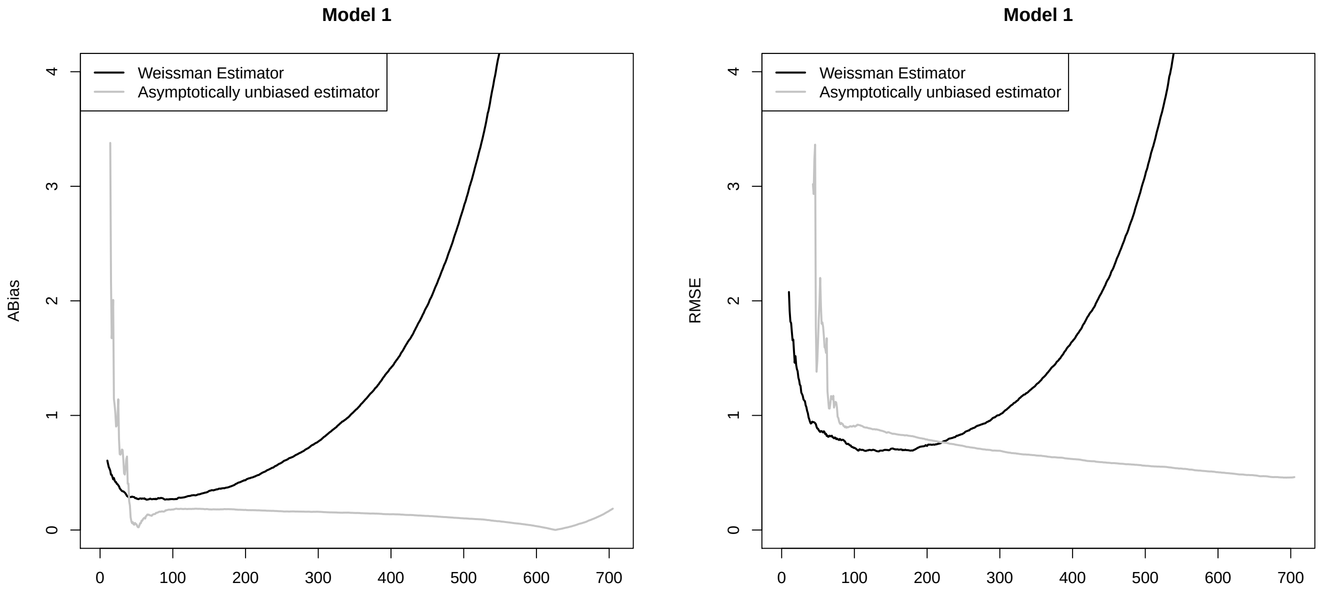
Figure 5: **Estimating the extreme value index: Model 4.**



(a) ABias under Model 4.

(b) RMSE under Model 4.

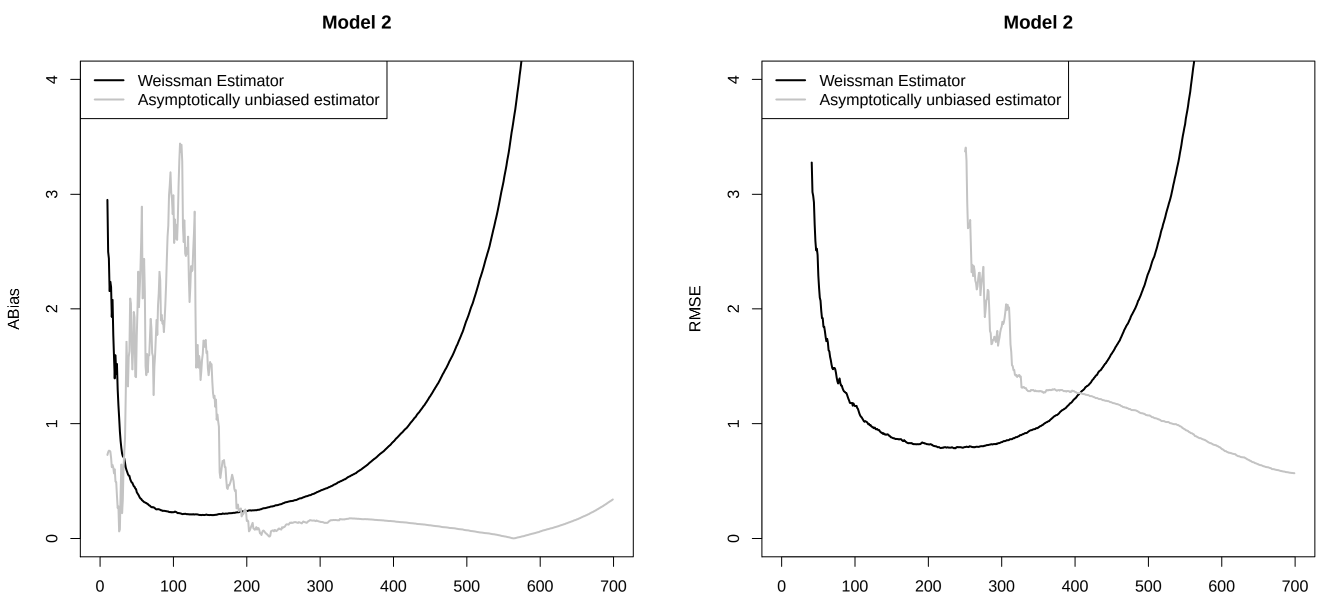
Figure 6: **Estimating the high quantile: Model 1.**



(a) ABias under Model 1.

(b) RMSE under Model 1.

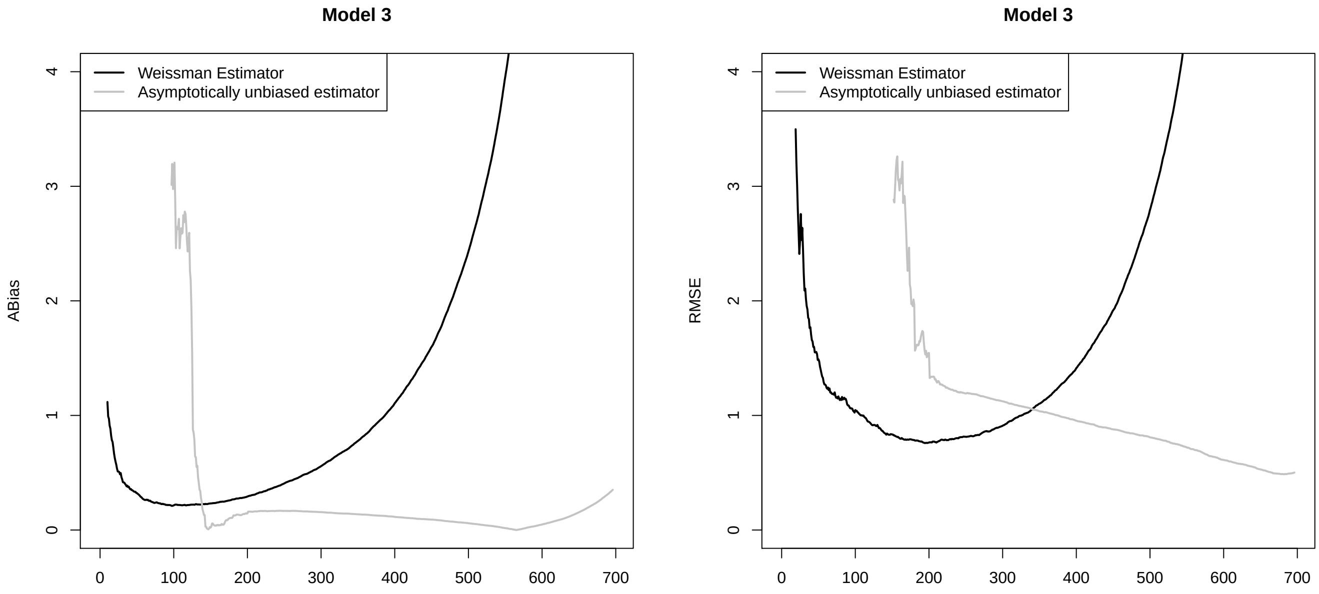
Figure 7: **Estimating the high quantile: Model 2.**



(a) ABias under Model 2.

(b) RMSE under Model 2.

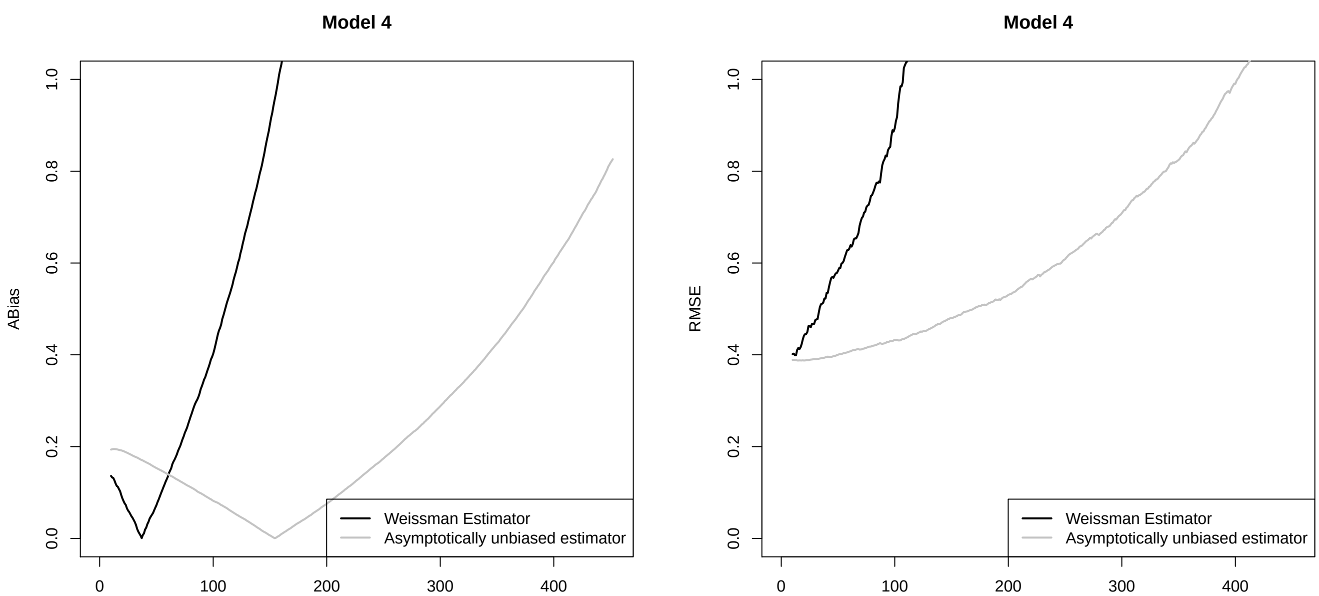
Figure 8: **Estimating the high quantile: Model 3.**



(a) ABias under Model 3.

(b) RMSE under Model 3.

Figure 9: **Estimating the high quantile: Model 4.**

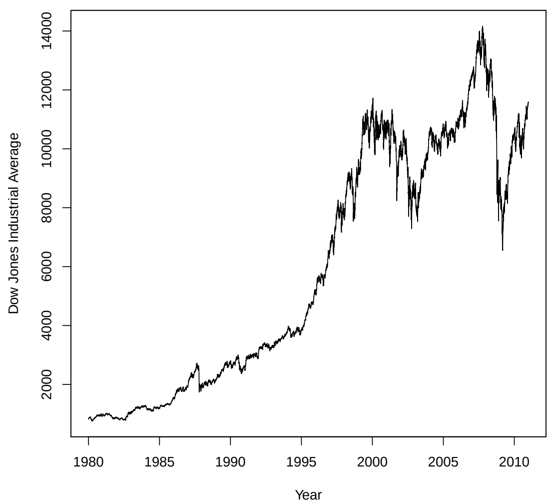


(a) ABias under Model 4.

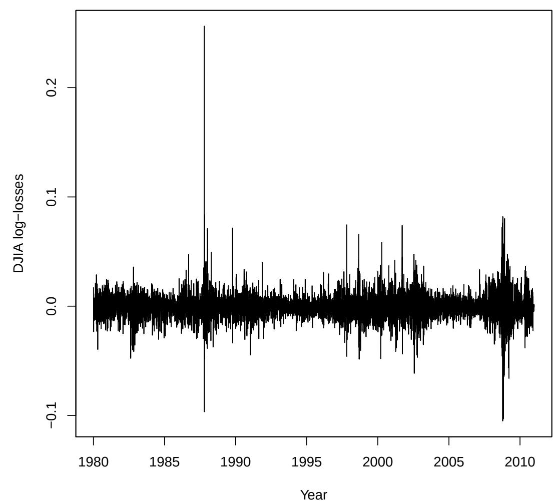
(b) RMSE under Model 4.

daily index from 1980 to 2010 and compute the daily loss returns. The indices and loss returns are presented in Figure 10a and 10b. From the figures, we observe that although the loss return series can be regarded as stationary, there is evidence of serial dependence such as volatility clustering. More concretely, by fitting the GARCH(1,1) model with Student-t distributed innovations to our dataset, we obtain estimates as $\hat{\lambda}_0 = 8.26 \times 10^{-7}$, $\hat{\lambda}_1 = 0.052$, $\hat{\lambda}_2 = 0.941$ and $\hat{\nu} = 5.64$. The existence of serial dependence prevents us from treating the series as i.i.d. observations. The serial dependence has to be accounted for when performing extreme value analysis.

Figure 10: **Historical time series of the DJIA index.** The figures in the left and right panels show the daily prices and loss returns of the DJIA index from 1980 to 2010 respectively.



(a) Dow Jones Industrial Average (DJIA).



(b) Loss returns of DJIA.

Our goal is to estimate the Value-at-Risk of the return series at 99.9% level, which corresponds to a high quantile with tail probability 0.1%, i.e. $x(0.001)$. From 8088 daily observations, a non-parametric estimate can be obtained by taking the eighth highest order statistic. We thus get 7.16% as the empirical estimate.

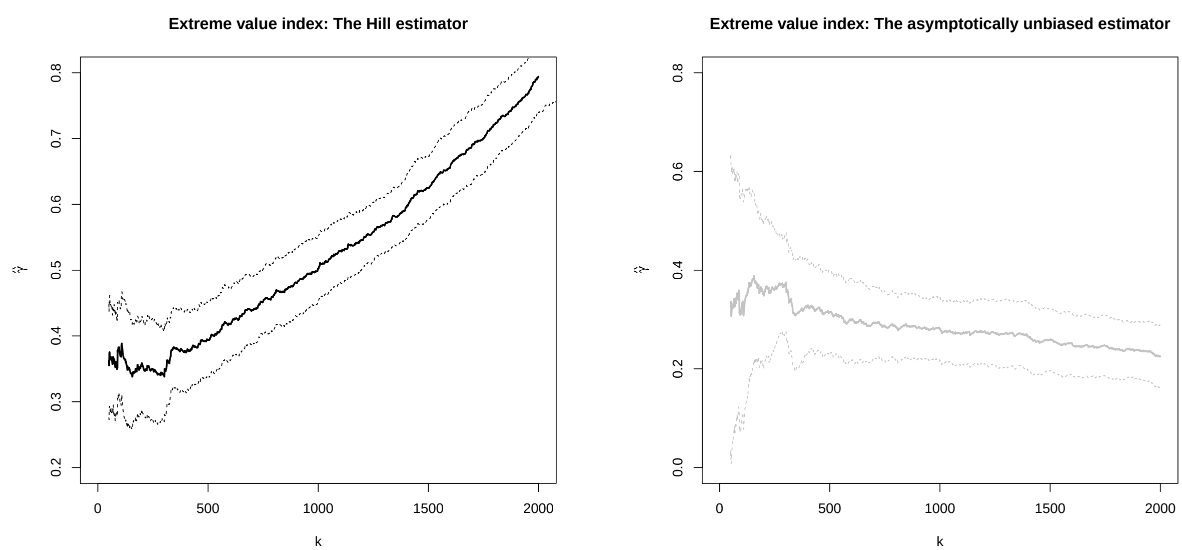
Next, we apply both the original Hill estimator and the asymptotically unbiased estimator to estimate the extreme value index of the loss return series. We start with estimating the second order parameter ρ. Following the estimation procedure in Section 6.2, we choose $k_\rho = 3515$ and obtain that $\hat{\rho} = -0.611$. Next we apply both estimators for $k_n = 50, 11, \cdots, 2000$. Since we do not employ a parametric model for the time series, there is no explicit formula for calculating the asymptotic variance of the two estimators. Therefore, we opt to use a block bootstrapping method to construct the confidence interval for the extreme value index.

The block bootstrapping follows the routine `tsboot` in the package `boot` in R. The block lengths are choosing to have a geometric distribution (`sim=geom`) with mean `l=200`. By repeating such a bootstrapping procedure 50 times, we obtain 50 bootstrapped estimates for each estimator. The sample standard deviation across the 50 estimates gives an estimate of the standard deviation of the underlying estimator for given k_n. We construct the 95% confidence interval using the point estimate and the estimated standard deviation. This procedure is applied to all values of k_n and for both estimators. The point estimates of the extreme value index as well as the lower and upper bounds of the confidence intervals are plotted against different choices of k_n in Figure 11.

Lastly, we apply both the original Weissman estimator and the asymptotically unbiased version to estimate the VaR at 99.9% level. The construction of the confidence intervals follows a similar block bootstrapping procedure. The results are plotted in Figure 12.

From the two figures, we observe that the estimates using the bias correction technique stays stable for a larger range of k values. In contrast, the estimates based on the original Hill estimator suffers from a large bias starting from $k \geq 400$. When applying the original EVT estimators, it is possible to choose k only around 250, which corresponds to 3% of the total sample. Correspondingly, we obtain an estimated extreme value index at 0.349 from the Hill estimator and an estimated VaR at 0.06549 from the Weissman estimator. With our asymptotically unbiased estimators, we obtain can take $k = 1000$ and obtain an estimated extreme value index at 0.280 with an estimated VaR at 0.05898. Note that the point estimates of the VaR are below, but close to, the empirical estimate. In addition to the point estimation, we investigate the confidence intervals of the estimated VaR. The Weissman estimator results in a 95% confidence interval as $[0.04268, 0.08831]$, while the confidence interval obtained from our asymptotically unbiased estimator is $[0.04219, 0.07577]$. Hence, we conclude that the bias correction procedure helps to obtain a more accurate estimate with a narrower confidence interval.

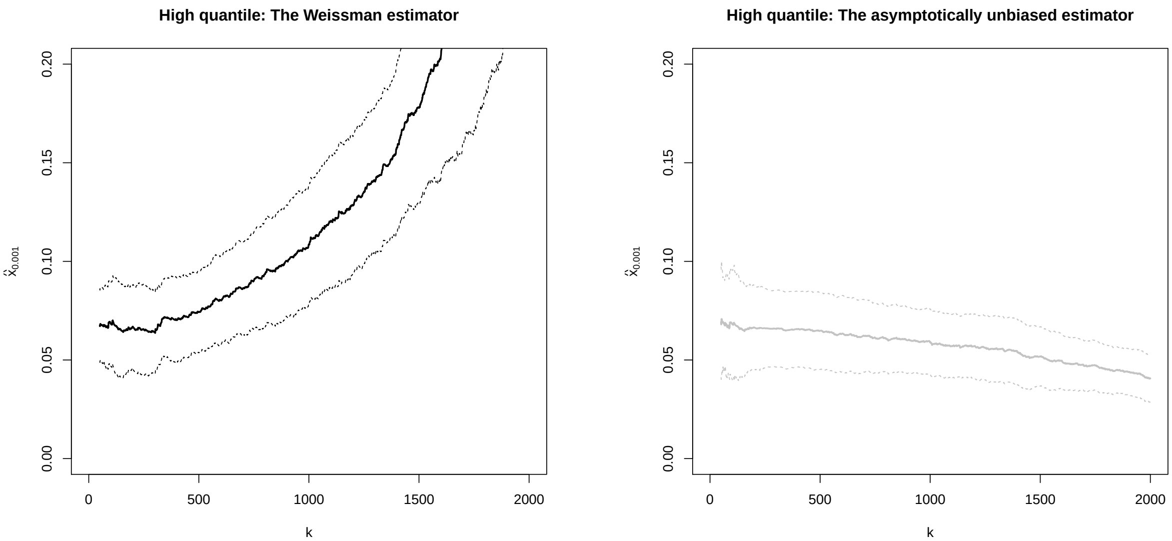
Figure 11: **Estimating the extreme value index for the DJIA index.** The figures present the estimates of the extreme value index for the loss returns of the DJIA index with varying choice of k. The left panel uses the Hill estimator. The right panel uses the asymptotically unbiased estimators of the extreme value index in Section 4.2. The confidence interval is obtained from the block bootstrapping method.



(a) The Hill estimator

(b) The asymptotically unbiased estimator

Figure 12: **Estimating the VaR at** 99.9% **level for the DJIA index.** The figures present the estimates of the VaR at 99.9% level for the loss returns of the DJIA index with varying choice of k. The left panel uses the Weissman estimator. The right panel uses the asymptotically unbiased estimators of high quantile provided in Section 4.3. The confidence interval is obtained from the block bootstrapping method.



(a) The Weissman estimator (b) The asymptotically unbiased estimator

A – Appendix – Proofs

The asymptotically unbiased estimator of the extreme value index is based on the moments

$$M_k^{(\alpha)} := \frac{1}{k}\sum_{i=1}^{k}\left(\log X_{n-i+1,n} - \log X_{n-k,n}\right)^{\alpha} \ ,$$

defined in Subsection 4.1. One can write these statistics as functionals of the tail quantile process $\left\{Q_n(t) := X_{n-[kt],n}\right\}_{t\in[0,1]}$ as follows

$$M_k^{(\alpha)} = \int_0^1 \left(\log \frac{Q_n(t)}{Q_n(1)}\right)^{\alpha} dt \ .$$

Therefore, to derive the asymptotic property of the asymptotically unbiased estimator, we first establish those of the tail quantile process and the moments. We first show that the tail quantile process can be approximated by a Gaussian process as in the following proposition.

Proposition A.1. *Suppose that $\{X_1, X_2, \dots\}$ is a stationary β-mixing time series with continuous common marginal distribution function F. Assume that F satisfies the third order condition (8) with parameters $\gamma > 0$, $\rho < 0$ and $\rho' \le 0$. Suppose that an intermediate sequence k satisfies that as $n \to \infty$, $k \to \infty$, $k/n \to 0$ and $\sqrt{k}A(n/k)B(n/k) = O(1)$. In addition, assume that the regulatory conditions holds. Then, for a given $\varepsilon > 0$, under a Skorohod construction, there exists two functions $\tilde{A} \sim A$ and $\tilde{B} = O(B)$, where A and B are the second and third order scale functions in (8),* *and a centered Gaussian process* $\{e(t)\}_{t\in[0,1]}$ *with covariance function* r *defined as in the regulatory condition (b), such that, as* $n\to\infty$,

$$\sup_{t\in(0,1]} t^{1/2+\varepsilon}\left|\sqrt{k}\left(\log\left(\frac{Q_n(t)}{U(n/k)}\right)+\gamma\log(t)\right)-\gamma t^{-1}e(t)\right.$$
$$\left.-\sqrt{k}\tilde{A}(n/k)\frac{t^{-\rho}-1}{\rho}-\sqrt{k}\tilde{A}(n/k)\tilde{B}(n/k)\frac{t^{-\rho-\rho'}-1}{\rho+\rho'}\right|\to 0 \ a.s.$$

Proof of Proposition A.1. By writing $X_i=U(Y_i)$ where each Y_i follows a standard Pareto distribution, we obtain that $\{Y_1,Y_2,\cdots\}$ is a stationary $\beta-$mixing series satisfying the regulatory conditions. This is a direct consequence of $Y_i=1/(1-F(X_i))$. We write $Q_n(t)=X_{n-[kt],n}=U(Y_{n-[kt],n})$ and focus first on the asymptotic property of the process $\left\{Y_{n-[kt],n}\right\}_{t\in[0,1]}$. By verifying the conditions in Drees [2003, Theorem 2.1], we get that under a Skorohod construction, there exists a centered Gaussian process $\{e(t)\}_{t\in[0,1]}$ with covariance function r defined in the regulatory condition (b), such that for $\varepsilon>0$, as $n\to\infty$,

$$\sup_{t\in(0,1]} t^{1/2+\varepsilon}\left|\sqrt{k}\left(t\frac{Y_{n-[kt],n}}{n/k}-1\right)-t^{-1}e(t)\right|\to 0 \text{ a.s..} \tag{16}$$

Next, we present an inequality on the U function based on the third order condition (8). Under the third order condition, there exists two functions $\tilde{A}\sim A$ and $\tilde{B}=O(B)$, such that for any $\delta>0$, there exists some positive number $u_0=u_0(\varepsilon)$ such that for all $u\geq u_0$ and $ux\geq u_0$,

$$\left|\frac{\frac{\log U(ux)-\log U(u)-\gamma\log x}{\tilde{A}(u)}-\frac{x^\rho-1}{\rho}}{\tilde{B}(u)}-\frac{x^{\rho+\rho'}-1}{\rho+\rho'}\right|\leq \delta x^{\rho+\rho'}\max(x^\delta,x^{-\delta})\,. \tag{17}$$

This inequality is a direct consequence of applying de Haan and Ferreira [2006, Theorem B.3.10] to the function $f(u)=\log U(u)-\gamma\log u$.

We combine the asymptotic property of $\left\{Y_{n-[kt],n}\right\}_{t\in[0,1]}$ in (16) with the inequality (17) as follows. Taking $u=n/k$ and $ux=Y_{n-[kt],n}$ in (17), we get that given any $0<\delta<-\rho-\rho'$, for sufficiently large $n>n_0(\delta)$, with probability 1,

$$\left|\log Q_n(t)-\log U(n/k)-\gamma\log\left(\frac{k}{n}Y_{n-[kt],n}\right)-\tilde{A}(n/k)\frac{\left(\frac{k}{n}Y_{n-[kt],n}\right)^\rho-1}{\rho}\right.$$
$$\left.-\tilde{A}(n/k)\tilde{B}(n/k)\frac{\left(\frac{k}{n}Y_{n-[kt],n}\right)^{\rho+\rho'}-1}{\rho+\rho'}\right|\leq\delta\tilde{A}(n/k)\tilde{B}(n/k)\left(\frac{k}{n}Y_{n-[kt],n}\right)^{\rho+\rho'+\delta}\,. \tag{18}$$

By applying (16), we bound the four terms in (18) that contain $\frac{k_n}{n}Y_{n-[k_n t],n}$ as

$$t^{1/2+\varepsilon}\left|\sqrt{k}\left(\log\left(\frac{k}{n}Y_{n-[kt],n}\right)+\log t\right)-t^{-1}e(t)\right| \to 0 \text{ a.s.},$$

$$t^{1/2+\varepsilon}\left|\sqrt{k}\left(\frac{\left(\frac{k}{n}Y_{n-[kt],n}\right)^{\rho}-1}{\rho}-\frac{t^{-\rho}-1}{\rho}\right)-t^{-\rho-1}e(t)\right| = o(t^{-\rho}) \to 0 \text{ a.s.},$$

$$t^{1/2+\varepsilon}\left|\sqrt{k}\left(\left(\frac{k}{n}Y_{n-[kt],n}\right)^{\rho+\rho'}-t^{-\rho-\rho'}\right)-(\rho+\rho')\left(t^{-\rho-\rho'-1}e(t)\right)\right| = o\left(t^{-\rho-\rho'}\right) \to 0 \text{ a.s.},$$

$$t^{1/2+\varepsilon}\left(\frac{k}{n}Y_{n-[kt],n}\right)^{\rho+\rho'+\delta} = O(t^{1/2-\rho-\rho'+\varepsilon-\delta}) = O(1) \text{ a.s.}$$

When taking $n\to\infty$, with the facts that $\sup_{t\in(0,1]} t^{1/2+\varepsilon}t^{-1}|e(t)| = O(1)$ a.s., $\sqrt{k}\tilde{A}(n/k)\tilde{B}(n/k) = O(1)$ and $\tilde{A}(n/k), \tilde{B}(n/k)\to 0$, the proposition is proved due to the free choice of δ. $\square$

By applying Proposition A.1, we get the asymptotic property of the moments $M_k^{(\alpha)}$ as follows.

Corollary A.2. *Assume that the conditions in Proposition A.1 hold. Then, under the same Skorokhod construction as in Proposition A.1, as $n\to\infty$*

$$\begin{aligned}&\sqrt{k}\left(M_k^{(\alpha)}-\gamma^{\alpha}\Gamma(\alpha+1)\right)-\alpha\gamma^{\alpha}P_1^{(\alpha)}-\sqrt{k}\tilde{A}(n/k)\gamma^{\alpha-1}\frac{\Gamma(\alpha+1)}{\rho}\left(\frac{1}{(1-\rho)^{\alpha}}-1\right)\\&-\sqrt{k}\tilde{A}(n/k)\tilde{B}(n/k)\gamma^{\alpha-1}\frac{\Gamma(\alpha+1)}{\rho+\rho'}\left(\frac{1}{(1-\rho-\rho')^{\alpha}}-1\right)\\&-\sqrt{k}\tilde{A}(n/k)^2\gamma^{\alpha-2}\frac{\Gamma(\alpha+1)}{2\rho^2}\left(\frac{1}{(1-2\rho)^{\alpha}}-\frac{2}{(1-\rho)^{\alpha}}+1\right)\to 0 \text{ a.s.,}\end{aligned}$$

where $P_1^{(\alpha)}$ are normally distributed random variables with mean zero. In addition

$$\operatorname{Cov}(P_1^{(\alpha)},P_1^{(\tilde{\alpha})}) = \iint_{[0,1]^2}(-\log s)^{\alpha-1}(-\log t)^{\tilde{\alpha}-1}\left\{\frac{r(s,t)}{st}-\frac{r(s,1)}{s}-\frac{r(1,t)}{t}+r(1,1)\right\}ds\,dt\,,$$

with the covariance function r defined as in regulatory condition (b).

Proof of Corollary A.2. Recall that

$$M_k^{(\alpha)} = \int_0^1\left(\log\frac{Q_n(t)}{U(n/k)}-\log\frac{Q_n(1)}{U(n/k)}\right)^{\alpha}dt\,.$$

Under the same Skorokhod construction as in Proposition A.1, we get that as $n\to\infty$,

$$\begin{aligned}\sup_{t\in(0,1]} t^{1/2+\varepsilon}&\left|\sqrt{k}\left(\log\left(\frac{Q_n(t)}{Q_n(1)}\right)-\gamma(-\log t)\right)-\gamma(t^{-1}e(t)-e(1))-\sqrt{k}\tilde{A}(n/k)\frac{t^{-\rho}-1}{\rho}\right.\\&\left.-\sqrt{k}\tilde{A}(n/k)\tilde{B}(n/k)\frac{t^{-\rho-\rho'}-1}{\rho+\rho'}\right|\to 0 \text{ a.s.}\,.\end{aligned}$$

The second order expansion $(1+x)^\alpha = 1+\alpha x + \frac{\alpha(\alpha-1)}{2}x^2 + o(x^2)$ yields that, as $n\to\infty$,

$$\sup_{t\in(0,1]} t^{1/2+\varepsilon}\Bigg|\sqrt{k}\left(\left(\log\left(\frac{Q_n(t)}{Q_n(1)}\right)\right)^\alpha - \gamma^\alpha(-\log t)^\alpha\right) - \alpha\gamma^\alpha(-\log t)^{\alpha-1}(t^{-1}e(t)-e(1))$$
$$-\sqrt{k}\tilde{A}(n/k)\alpha\gamma^{\alpha-1}(-\log t)^{\alpha-1}\frac{t^{-\rho}-1}{\rho} - \sqrt{k}\tilde{A}(n/k)\tilde{B}(n/k)\alpha\gamma^{\alpha-1}(-\log t)^{\alpha-1}\frac{t^{-\rho-\rho'}-1}{\rho+\rho'}$$
$$-\sqrt{k}\tilde{A}^2(n/k)\frac{\alpha(\alpha-1)}{2}\gamma^{\alpha-2}(-\log t)^{\alpha-2}\left(\frac{t^{-\rho}-1}{\rho}\right)^2\Bigg| \to 0 \text{ a.s..}$$

Some terms are omitted because $\sup_{t\in(0,1]} t^{1/2+\varepsilon}t^{-1}|e(t)| = O(1)$ a.s. and $\tilde{A}(n/k)\to 0$ as $n\to\infty$.

By taking $\varepsilon < 1/2$, we can then take the integral of $\left(\log\left(\frac{Q_n(t)}{Q_n(1)}\right)\right)^\alpha$ on $(0,1]$ and use the fact that $\int_0^1(-\log t)^{a-1}t^{-b}dt = \dfrac{\Gamma(a)}{(1-b)^a}$ for $b<1$ to obtain the result in the corollary. The random term is $P_1^{(\alpha)} = \int_0^1(-\log t)^{\alpha-1}(t^{-1}e(t)-e(1))\,dt$. The covariance can be calculated from there. □

Next, we handle the estimator of the second order parameter ρ. The estimator of ρ is based on a different k sequence, k_ρ, satisfying (9). Because k_ρ satisfies the condition in Proposition A.1, we get the asymptotic properties of the moments $M_{k_\rho}^{(\alpha)}$ as in Corollary A.2. Then, following the same lines as in the proof of Gomes et al. [2002, Theorem 2.2], we get the following proposition.

Proposition A.3. *Suppose that $\{X_1, X_2, \ldots\}$ is a stationary β-mixing time series with continuous common marginal distribution function F. Assume that F satisfies the third order condition (8) with parameters $\gamma>0$, $\rho<0, \rho'\le 0$. Suppose that an intermediate sequence k_ρ satisfies (9). In addition, assume that the regulatory conditions holds. Then, for the ρ estimator defined in (11) and as $n\to\infty$*

$$\sqrt{k_\rho}\tilde{A}(n/k_\rho)\left(\hat{\rho}_{k_\rho}^{(\alpha)}-\rho\right)$$

is asymptotically normally distributed.

We remark that analogous to the result in Theorem 2.1 in Gomes et al. [2002], the consistency of the ρ estimator for β-mixing time series can be proved under only the second order condition (3) and weaker conditions on k_ρ.

Finally, we can use the tools built in Corollary A.2 and Proposition A.3 to prove our main results.

Proof of Theorem 4.1. From Corollary A.2, with k_n satisfying (10), under the same Skorokhod construction as in Proposition A.1, the Hill estimator has the following expansion

$$\sqrt{k_n}\left(\hat{\gamma}_{k_n}-\gamma\right) - \gamma P_1^{(1)} - \sqrt{k_n}\tilde{A}(n/k_n)\frac{1}{1-\rho} \to 0 \text{ a.s.}$$

which leads to

$$\sqrt{k_n}\left(\hat{\gamma}_{k_n}^2-\gamma^2\right) - 2\gamma^2 P_1^{(1)} - \sqrt{k_n}\tilde{A}(n/k_n)\frac{2\gamma}{1-\rho} \to 0 \text{ a.s. .}$$

Together with the asymptotic property of $M_{k_n}^{(2)}$ obtained again from Corollary A.2, it implies that

$$\sqrt{k_n}\left(M_{k_n}^{(2)} - 2\hat{\gamma}_{k_n}^2\right) - 2\gamma^2(P_1^{(2)} - 2P_1^{(1)}) - \sqrt{k_n}\tilde{A}(n/k_n)\frac{2\gamma\rho}{(1-\rho)^2} \to 0 \text{ a.s.} .$$

Thus, the asymptotic unbiased estimator has the following expansion, almost surely as $n \to \infty$,

$$\begin{aligned}
&\sqrt{k_n}\left(\hat{\gamma}_{k_n,k_\rho,\alpha} - \gamma\right) \\
=&\sqrt{k_n}\left(\hat{\gamma}_{k_n} - \gamma\right) - \frac{1}{2\hat{\gamma}_{k_n}\hat{\rho}_{k_\rho}^{(\alpha)}(1-\hat{\rho}_{k_\rho}^{(\alpha)})^{-1}}\sqrt{k_n}\left(M_{k_n}^{(2)} - 2\hat{\gamma}_{k_n}^2\right) \\
=&\gamma P_1^{(1)} + \sqrt{k_n}\tilde{A}(n/k_n)\frac{1}{1-\rho} - \frac{1}{2\hat{\gamma}_{k_n}\hat{\rho}_{k_\rho}^{(\alpha)}(1-\hat{\rho}_{k_\rho}^{(\alpha)})^{-1}}\left(2\gamma^2(P_1^{(2)} - 2P_1^{(1)}) + \sqrt{k_n}\tilde{A}(n/k_n)\frac{2\gamma\rho}{(1-\rho)^2}\right) \\
=&\gamma P_1^{(1)} - \frac{\gamma(1-\hat{\rho}_{k_\rho}^{(\alpha)})}{\hat{\rho}_{k_\rho}^{(\alpha)}}\left(P_1^{(2)} - 2P_1^{(1)}\right) + \sqrt{k_n}\tilde{A}(n/k_n)\frac{\rho}{(1-\rho)^2}\left(\frac{1-\rho}{\rho} - \frac{1-\hat{\rho}_{k_\rho}^{(\alpha)}}{\hat{\rho}_{k_\rho}^{(\alpha)}}\right).
\end{aligned} \tag{19}$$

In the last step we use the fact that $\hat{\gamma}_{k_n} \to \gamma$ a.s., as $n \to \infty$. Further, the relation $k_n/k_\rho \to 0$ implies that $\frac{\sqrt{k_n}\tilde{A}(n/k_n)}{\sqrt{k_\rho}\tilde{A}(n/k_\rho)} \to 0$, as $n \to \infty$. Thus, according to Proposition A.3 and Cramér's Delta method, we get that as $n \to \infty$,

$$\sqrt{k_n}\tilde{A}(n/k_n)\frac{\rho}{(1-\rho)^2}\left(\frac{1-\rho}{\rho} - \frac{1-\hat{\rho}_{k_\rho}^{(\alpha)}}{\hat{\rho}_{k_\rho}^{(\alpha)}}\right) \xrightarrow{\mathbb{P}} 0.$$

Together with the consistency of $\hat{\rho}_{k_\rho}^{(\alpha)}$, the expansion (19) implies that as $n \to \infty$,

$$\sqrt{k_n}\left(\hat{\gamma}_{k_n,k_\rho,\alpha} - \gamma\right) \xrightarrow{\mathbb{P}} \frac{\gamma}{\rho}\left(P_1^{(1)}(2-\rho) + P_1^{(2)}(\rho - 1)\right).$$

The theorem is proved by using the covariance structure of $\left(P_1^{(1)}, P_1^{(2)}\right)$ given in Corollary A.2. $\square$

Proof of Theorem 4.2. Denote $d_n := k_n/(np_n)$ and $T_n = \dfrac{(M_{k_n}^{(2)} - 2\hat{\gamma}_{k_n}^2)(1-\hat{\rho}_{k_\rho}^{(\alpha)})^2}{2\hat{\gamma}_{k_n}\{\hat{\rho}_{k_\rho}^{(\alpha)}\}^2}$.

With $P_1^{(\alpha)}$ defined in Corollary A.2, following the lines in the proof of Theorem 4.1, we obtain that, under the same Skorokhod construction as in Proposition A.1,

$$\sqrt{k_n}\left(T_n - \frac{\tilde{A}\left(\frac{n}{k_n}\right)}{\rho}\right) - \frac{\gamma(1-\rho)^2}{\rho^2}(P_1^{(2)} - 2P_1^{(1)}) \to 0 \quad \text{a.s.}, \tag{20}$$

as $n \to \infty$, which implies that $T_n \to 0$ *a.s.* Together with $\sqrt{k_n}\tilde{A}^2\left(\frac{n}{k_n}\right) \to 0$ as required in condition (10) we have a stronger result that is, as $n \to \infty$,

$$\sqrt{k_n}\tilde{A}\left(\frac{n}{k_n}\right)T_n \to 0 \; a.s. . \tag{21}$$

Consider the following expansion

$$
\begin{aligned}
&\frac{\sqrt{k_n}}{\log d_n}\left(\frac{\hat{x}_{k_n,k_\rho,\alpha}(p_n)}{x\left(p_n\right)}-1\right)\\
=&\frac{\sqrt{k_n}}{\log d_n}\left(\frac{X_{n-k_n,n}d_n^{\hat{\gamma}_{k_n,k_\rho,\alpha}}}{x\left(p_n\right)}-1\right)(1-T_n)-\frac{\sqrt{k_n}}{\log d_n}T_n\\
=&\frac{d_n^{\gamma}U\left(\frac{n}{k_n}\right)}{U\left(\frac{1}{p_n}\right)}\left[\frac{\sqrt{k_n}}{\log d_n}\left(\frac{X_{n-k_n,n}}{U\left(\frac{n}{k_n}\right)}-1\right)d_n^{\hat{\gamma}_{k_n,k_\rho,\alpha}-\gamma}+\frac{\sqrt{k_n}}{\log d_n}\left(d_n^{\hat{\gamma}_{k_n,k_\rho,\alpha}-\gamma}-1\right)\right]\cdot(1-T_n)\\
&-\frac{\sqrt{k_n}}{\log d_n}\left(T_n-\frac{\tilde{A}\left(\frac{n}{k_n}\right)}{\rho}\right)+T_n\cdot\frac{\sqrt{k_n}\tilde{A}\left(\frac{n}{k_n}\right)}{\log d_n}\frac{\frac{U\left(\frac{1}{p_n}\right)d_n^{-\gamma}}{U\left(\frac{n}{k_n}\right)}-1}{\tilde{A}\left(\frac{n}{k_n}\right)}\frac{d_n^{\gamma}U\left(\frac{n}{k_n}\right)}{U\left(\frac{1}{p_n}\right)}\\
&-\frac{\sqrt{k_n}\tilde{A}^2\left(\frac{n}{k_n}\right)}{\log d_n}\frac{\frac{U\left(\frac{1}{p_n}\right)d_n^{-\gamma}}{U\left(\frac{n}{k_n}\right)}-1}{\tilde{A}\left(\frac{n}{k_n}\right)}\frac{\frac{d_n^{\gamma}U\left(\frac{n}{k_n}\right)}{U\left(\frac{1}{p_n}\right)}-1}{\tilde{A}\left(\frac{n}{k_n}\right)}\\
&-\frac{\sqrt{k_n}\tilde{A}\left(\frac{n}{k_n}\right)\left(\tilde{A}\left(\frac{n}{k_n}\right)+\tilde{B}\left(\frac{n}{k_n}\right)\right)}{\log d_n}\frac{\frac{\frac{U\left(\frac{1}{p_n}\right)d_n^{-\gamma}}{U\left(\frac{n}{k_n}\right)}-1}{\tilde{A}\left(\frac{n}{k_n}\right)}+\frac{1}{\rho}}{\tilde{A}\left(\frac{n}{k_n}\right)+\tilde{B}\left(\frac{n}{k_n}\right)}\\
=:&I_1-I_2+I_3-I_4-I_5.
\end{aligned}
$$

The third order condition in (8) implies that as $n\to\infty$,

$$
\left|\frac{\frac{U\left(\frac{1}{p_n}\right)d_n^{-\gamma}}{U\left(\frac{n}{k_n}\right)}-1}{\tilde{A}\left(\frac{n}{k_n}\right)}+\frac{1}{\rho}\right|=O\left(\tilde{A}\left(\frac{n}{k_n}\right)+\tilde{B}\left(\frac{n}{k_n}\right)\right). \tag{22}
$$

The limit relation in (22) further implies that as $n\to\infty$,

$$
\frac{\frac{U\left(\frac{1}{p_n}\right)d_n^{-\gamma}}{U\left(\frac{n}{k_n}\right)}-1}{\tilde{A}\left(\frac{n}{k_n}\right)}\to-\frac{1}{\rho}\qquad\text{and}\qquad\frac{U\left(\frac{1}{p_n}\right)d_n^{-\gamma}}{U\left(\frac{n}{k_n}\right)}\to 1.
$$

Combining (22) with condition (10), we get that $I_4\to 0$ and $I_5\to 0$ as $n\to\infty$. Next, from (21), we get that $I_2\to 0$ and $I_3\to 0$ a.s., as $n\to\infty$.

Lastly, we deal with the term I_1. Denote the limit of $\sqrt{k_n}\left(\hat{\gamma}_{k_n,k_\rho,\alpha}-\gamma\right)$ as Γ. Then we have that as $n\to\infty$

$$
\frac{\sqrt{k_n}}{\log d_n}\left(d_n^{\hat{\gamma}_{k_n,k_\rho,\alpha}-\gamma}-1\right)\to\Gamma,\quad\text{a.s.,}
$$

which implies that $\frac{1}{\log d_n} d_n^{\hat{\gamma}_{k_n,k_\rho,\alpha}-\gamma} \to 0$ a.s. Together with the facts that $T_n \to 0$ a.s. and

$$\sqrt{k_n}\left(\frac{X_{n-k_n,n}}{U\left(\frac{n}{k_n}\right)}-1\right)=O(1) \text{ a.s.}$$

as $n \to \infty$, we get that $I_1 \to \Gamma$ a.s. as $n \to \infty$. The theorem is proved by combining the limit properties of the five terms in the expansion. □

Acknowledgments

The authors would like to thank two anonymous referees and the associate editor for their helpful comments. Laurens de Haan's research was partially supported by ENES project EXPL/MAT-STA/0622/2013. Cécile Mercadier and Chen Zhou's research was partially supported by the "EUR fellowship" grant.

References

P. Billingsley. *Probability and measure.* John Wiley & Sons, 1979.

J.-J. Cai, L. de Haan, and C. Zhou. Bias correction in extreme value statistics with index around zero. *Extremes*, 16(2):173–201, 2013.

G. Ciuperca and C. Mercadier. Semi-parametric estimation for heavy tailed distributions. *Extremes*, 13: 55–87, 2010.

J. Danielsson, L. de Haan, L. Peng, and C. G. de Vries. Using a bootstrap method to choose the sample fraction in tail index estimation. *J. Multivariate Anal.*, 76(2):226–248, 2001.

L. de Haan and A. Ferreira. *Extreme value theory. An introduction.* Springer Series in Operations Research and Financial Engineering. New York, NY: Springer., 2006.

F.X. Diebold, T. Schuermann, and J.D. Stroughair. Pitfalls and opportunities in the use of extreme value theory in risk management. *Journal of Risk Finance, The*, 1(2):30–35, 2000.

H. Drees. Weighted approximations of tail processes for β-mixing random variables. *Ann. Appl. Probab.*, 10 (4):1274–1301, 2000.

H. Drees. Tail empirical processes under mixing conditions. In *Empirical process techniques for dependent data*, pages 325–342. Birkhäuser Boston, Boston, MA, 2002.

H. Drees. Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli*, 9(4): 617–657, 2003.

Holger Drees and Edgar Kaufmann. Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Process. Appl.*, 75(2):149–172, 1998. ISSN 0304-4149. doi: 10.1016/S0304-4149(98) 00017-9. URL http://dx.doi.org/10.1016/S0304-4149(98)00017-9.

C. M Goldie. Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.*, 1(1): 126–166, 1991.

I.M. Gomes, L. de Haan, and L.H. Rodrigues. Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):31–52, 2008.

M. I. Gomes, L. de Haan, and L. Peng. Semi-parametric estimation of the second order parameter in statistics of extremes. *Extremes*, 5(4):387–414, 2002.

Armelle Guillou and Peter Hall. A diagnostic for selecting the threshold in extreme value analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 63(2):293–305, 2001. ISSN 1369-7412. doi: 10.1111/1467-9868.00286. URL http://dx.doi.org/10.1111/1467-9868.00286.

B.M. Hill. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3(5): 1163–1174, 1975.

T. Hsing. On tail index estimation using dependent data. *Ann. Statist.*, 19(3):1547–1569, 1991.

H. Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, 131(1):207–248, 1973.

A. J. McNeil and R. Frey. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance*, 7(3):271–300, 2000.

S. Resnick and C. Stărică. Asymptotic behavior of Hill's estimator for autoregressive data. *Comm. Statist. Stochastic Models*, 13(4):703–721, 1997.

H. Rootzén. The tail empirical process for stationary sequence. Preprint, Chalmers Univ. Gothenburg, 1995.

H. Rootzén. Weak convergence of the tail empirical process for dependent sequences. *Stochastic Process. Appl.*, 119(2):468–490, 2009.

C. Stărică. On the tail empirical process of solutions of stochastic difference equations. Preprint, Chalmers Univ. Gothenburg, 1999. URL www.math.chalmers.se/-starica/resume/aarch.ps.gz.

P. Sun and C. Zhou. Diagnosing the distribution of garch innovations. *Unpublished manuscript*, 2013.

I. Weissman. Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.*, 73(364):812–815, 1978.